Probabilistic Linkage
of Vital Event Records in Scotland
using Familial Groups

Özgür Akgün, Tom Dalton, Alan Dearle, Eilidh Garrett, Graham Kirby

The Workshop for the Systematic Linking of Historical Records

University of Guelph, Canada – May 11-13, 2017

Probabilistic Linkage of Vital Event Records in Scotland using Familial Groups

- Context: ADRN, ADRC-S, Digitising Scotland
- Linkage: The Standard Way
- This work is exploratory
- Key Points
  - Forming Conjugal Family Units (CFUs)
    - instead of linking individuals
  - Using high-dimensional similarity search (M-Tree)
    - instead of blocking
- Results

Digitising Scotland

- Digitise the 24 million Scottish vital events record images (births, marriages and deaths) since 1855.

- This will allow research access to individual-level information on some 18 million individuals - a large proportion of those who have lived in Scotland since 1855.

- Establish links between related records to determine genealogical structures.


Norwegian Sea
Greenland
Iceland
Sweden
Finland
Norway
Russia
Canada
Hudson Bay
United Kingdom
Ireland
Denmark
Belarus
Poland
Germany
Ukraine
France
Kazakhstan
Mongolia
Romania
Italy
Spain
Portugal
Greece
Turkey
Uzbekistan
Turkmenistan
China
South Korea
North Atlantic Ocean
Iraq
Iran
Afghanistan
Pakistan
Morocco
Algeria
Libya
Egypt
Saudi Arabia
India
Nepal
Myanmar (Burma)
Thailand
Vietnam
Philippines
Malaysia
Indonesia
Mauritania
Mali
Niger
Chad
Sudan
Yemen
Oman
Nigeria
South Sudan
Ethiopia
Somalia
Kenya
DR Congo
Tanzania
Angola
Zambia
Mozambique
Namibia
Zimbabwe
Botswana
Madagascar
South Africa
Indian Ocean
Australia
United States
Mexico
Gulf of Mexico
Cuba
Caribbean Sea
Nicaragua
Venezuela
Colombia
Brazil
Peru
Bolivia
Paraguay
Chile
Argentina
South Atlantic Ocean


Norwegian Sea
Iceland
Sweden
Norway
Finland
Estonia
Latvia
Lithuania
Baltic Sea
North Sea
Denmark
United Kingdom
Ireland
Netherlands
Poland
Belarus
Germany
Belgium
Czechia
Slovakia
Ukraine
Moldova
Hungary
France
Romania
Croatia
Serbia
Italy
Bulgaria
Portugal
Spain
Greece
Turkey
Georgia
Azerbaijan
Kazakhstan
Uzbekistan
Turkmenistan
Kyrgyzstan
Tajikistan
Afghanistan
Iran
Iraq
Syria
Lebanon
Israel
Jordan
Pakistan
Nepal
Tunisia
Morocco
Algeria
Libya
Egypt
North Atlantic Ocean
Mediterranean Sea
Black Sea
Caspian Sea
Tyrrhenian Sea
Moscow
Москва
Astana
Астана
Riyadh
الرياض
London
Paris
Berlin
Prague
Vienna
Rome
Madrid
Barcelona
Istanbul
Western


Iceland
SCOTLAND
Inverness
Aberdeen
Dundee
Edinburgh
Glasgow
United Kingdom
NORTHERN IRELAND
Isle of Man
Liverpool
Manchester
Dublin
Galway
Ireland
Limerick
Cork
ENGLAND
Cambridge
WALES
Oxford
Bristol
Cardiff
London
Brighton
Southampton
Plymouth
North Atlantic Ocean
Kazakhstan
Uzbekistan
Kyrgyzstan
Turkmenistan
Tajikistan
Afghanistan
Pakistan
New Delhi
Nepal

Probabilistic Data Linkage

- Task: Find records that refer to the same entity across different data sources
- Abundance of data
- Lack of unique IDs (in general)
- Enables
  - Answering some interesting questions
  - Potentially effecting policy decisions
  - Direct/Indirect cost savings
    - (direct: avoiding a full census)

Probabilistic linkage algorithms: How do they work

- Similarity metric between pairs of individuals on records
- Compare candidate pairs
- Threshold for definite-links
- Threshold for definite-nonlinks
- Everything in the middle: clerical review
- Feedback
  - In the form of adjusting the similarity metric or thresholds
- Done!

Problems

- Similarity metric
- Too many pairwise comparisons
- Thresholds
- Evaluation
- Historical data

Data for this experiment

- We use two manually-linked data sets
  - Isle of Skye
    - 17613 birth records
    - 12285 death records
    - 2669 marriage records
  - Kilmarnock
    - 38431 birth records
    - 23715 death records
    - 8654 marriage records

- Digitising Scotland: all the vital events recorded in Scotland between 1855 and 1977

Data for this experiment

data sets

al events recorded in
1977

Data for this experiment

KILMARNOCK

data

al ev

1977


Edinbane
Dunvegan
Portree
RAASAY
Duirinish
WIAY
1
2
3
Skye
Bridge
Cuillin
Hills
Kyleakin
Broadford
Minginish
SOAY
Strathaird
CANNA
RÙM
Sleat

al ev
1977

Birth records

1. Unique 'Record' Identifier
2. Child’s surname
3. Child’s Forename
4. Sex
5. Year of Registration
6. Registration District Number
7. Registration District Suffix
8. Entry
9. Birth Day
10. Birth Month
11. Birth Year
12. Mother’s Maiden Surname
13. Birth Address
14. Father’s Forename
15. Father’s Surname
16. Father’s Occupation
17. Mother’s Forename
18. Mother’s Surname
19. Parent’s Day of Marriage
20. Parent’s Month of Marriage
21. Parent’s Year of Marriage
22. Parents Place of Marriage
23. Illegitimate indicator
24. Informant
25. Informant did not Sign
26. Adoption

Birth records

1. Unique 'Record' Identifier
2. Child's surname
3. Child's Forename
4. Sex
5. Year of Registration
6. Registration District Number
7. Registration District Suffix
8. Entry
9. Birth Day
10. Birth Month
11. Birth Year
12. Mother's Maiden Surname
13. Birth Address
14. Father's Forename
15. Father's Surname
16. Father's Occupation
17. Mother's Forename
18. Mother's Surname
19. Parent's Day of Marriage
20. Parent's Month of Marriage
21. Parent's Year of Marriage
22. Parents Place of Marriage
23. Illegitimate indicator
24. Informant
25. Informant did not Sign
26. Adoption

BIRTHS in the District (1.) of Wilton (2.) in the County (3.) of Roxburgh (4.) (5.) (6.)

No.	Name and Surname	When and Where Born.	Sex.	Name, Surname, & Rank or Profession of Father. Name, and Maiden Surname of Mother. Date and Place of Marriage.	Signature and Qualification of Informant, and Residence, if out of the House in which the Birth occurred.	When and Where Registered and Signature of Registrar.
10	Frank Rutherford McCartney	1895. January Seventh. 6h. 15m. P.M. 4 Howlandstterrace Wilton	M.	James McCartney, Frame-work-Knitter. Agnes McCartney, N. S. Pirnie. 1876 Nov. 9th North Leith	James McCartney Father (Present)	1895. January At Wilton James Shiel
11	Helen McCartney	1895. January Seventh. 6h. 20m. P.M. 4 Howlandstterrace Wilton	F.	James McCartney, Frame-work-Knitter. Agnes McCartney, N. S. Pirnie. 1876 Nov. 9th North Leith	James McCartney Father. (Present)	1895. January At Wilton James Shiel,
12	Andrew Blyth Oliver	1895. January Twentieth 11h. 40m. P.M. 12 Wellington Street, Wilton	M.	Francis Oliver, Painter (Journeyman) Catherine Oliver, N. S. Welsh. 1894. April 27th Hawick	Francis Oliver Father (Present)	1895. January 2 At Wilton James Shiel

James Shiel, Regis

Death Records

1. Unique 'Record' Identifier
2. Surname
3. Forename
4. Sex
5. Year of Registration
6. Registration District Number
7. Registration District Suffix
8. Entry
9. Death Month
10. Death Day
11. Death Year
12. Age at Death
13. Date of Birth
14. Occupation
15. Marital Status
16. Spouse's Name(s)
17. Spouse's Occupation(s)
18. Place of Death
19. Father's Forename
20. Father's Surname
21. Father's Occupation
22. Father Deceased
23. Mother's Forename
24. Mother's Maiden Surname
25. Mother Deceased
26. Cause of Death
27. Certifying Doctor

Death Records

1. Unique 'Record' Identifier
2. Surname
3. Forename
4. Sex
5. Year of Registration
6. Registration District Number
7. Registration District Suffix
8. Entry
9. Death Month
10. Death Day
11. Death Year
12. Age at Death
13. Date of Birth
14. Occupation
15. Marital Status
16. Spouse's Name(s)
17. Spouse's Occupation(s)
18. Place of Death
19. Father's Forename
20. Father's Surname
21. Father's Occupation
22. Father Deceased
23. Mother's Forename
24. Mother's Maiden Surname
25. Mother Deceased
26. Cause of Death
27. Certifying Doctor

DEATHS in the District of Drainie in the County of Moray.

(1.) (2.) (3.) (4.) (5.) (6.) (7.) (8.)

No.	Name and Surname. Rank or Profession, and whether Single, Married, or Widowed.	When and Where Died.	Sex.	Age.	Name, Surname, and Rank or Profession of Father. Name, and Maiden Surname of Mother.	Cause of Death, Duration of Disease, and Medical Attendant by whom certified.	Signature and Qualification of Informant, and Residence, if out of the House in which the Death occurred.	When and where Registered and Signature of Registrar.
1	Margaret Mary Leslie	1930. January Ninth 5h. 0m. P.M. Sweethillock Drainie	F	3 months	William Watt Leslie Farm Servant. Margaret Helen Leslie M.S. Stuart	Chronic Enteritis Pneumonia Ag Certified by Gordon Thom M.B. Ch.B.	Isabella Stuart Aunt Kinkstile Farm Elgin 'Present'	1930. January 10 Ap Lossiemouth J.A.Peterkin Registrar
2	Jessie Mitchell Fishworker Single	1930. January Nineteenth 9h. 30m. P.M. 6 King Street Lossiemouth	F	70 Years	Robert Mitchell Fisherman (Deceased) Jessie Mitchell M.S. Stewart (Deceased)	Angina Pectoris Ag Certified by Thomas Brander M.B.C.M.	Joseph Main Nephew-in-law 1a Argyle Street Lossiemouth (Present)	1930. January 20 Ap Lossiemouth J.A.Peterkin Registrar
3	James Kirton Farmer (retired) Widower of 1st Barbara Minty 2nd Charlotte Minty	1930. January Thirtieth 9h. 45m. Am. 43 Maeduff Street Lossiemouth	M	75 Years	James Kirton Farmer (Deceased) Isabella Kirton M.S. Linn (Deceased)	Carcinoma Oesophagus and Stomach about 1 Year Secondary nodules Ag Certified by Alt Barclay L.R.C.S.&P.Ed.	William Thomson Son-in-law (Present)	1930. January 30 Ap Lossiemouth J.A.Peterkin Registrar

D

J.A.Peterkin Registrar.

Marriage records

1. Unique 'Record' Identifier
2. Groom’s Surname
3. Groom’s Forename
4. Bride Surname
5. Bride Forename
6. Marriage Day
7. Marriage Month
8. Marriage Year
9. Place of Marriage
10. Year of Registration
11. Registration District Number
12. Registration District Suffix
13. Entry
14. Denomination
15. Groom’s address
16. Groom age or date of birth
17. Groom’s occupation
18. Groom’s marital status
19. Bride’s address
20. Bride’s age or date of birth
21. Bride’s occupation
22. Bride’s marital status
23. Groom Father’s Forename
24. Groom Father’s Surname
25. Groom Father Deceased
26. Groom Mother’s Forename
27. Groom Mother’s Maiden Surname
28. Groom mother Deceased
29. Groom Father Occupation
30. Bride Father’s Forename
31. Bride Father’s Surname
32. Bride Father Deceased
33. Bride Mother’s Forename
34. Bride Mother’s Maiden Surname
35. Bride Mother Deceased
36. Bride Father Occupation
37. Groom did not Sign
38. Bride did not Sign

Marriage records

1. Unique 'Record' Identifier
2. Groom's Surname
3. Groom's Forename
4. Bride Surname
5. Bride Forename
6. Marriage Day
7. Marriage Month
8. Marriage Year
9. Place of Marriage
10. Year of Registration
11. Registration District Number
12. Registration District Suffix
13. Entry
14. Denomination
15. Groom's address
16. Groom age or date of birth
17. Groom's occupation
18. Groom's marital status
19. Bride's address
20. Bride's age or date of birth
21. Bride's occupation
22. Bride's marital status
23. Groom Father's Forename
24. Groom Father's Surname
25. Groom Father Deceased
26. Groom Mother's Forename
27. Groom Mother's Maiden Surname
28. Groom mother Deceased
29. Groom Father Occupation
30. Bride Father's Forename
31. Bride Father's Surname
32. Bride Father Deceased
33. Bride Mother's Forename
34. Bride Mother's Maiden Surname
35. Bride Mother Deceased
36. Bride Father Occupation
37. Groom did not Sign
38. Bride did not Sign

3. **MARRIAGES in the** District **of** Perth **in the** County **of** Perth.

No.	When, Where, and How Married.	Signatures of Parties. Rank or Profession, Whether Single or Widowed, and Relationship (if any).	Age.	Usual Residence.	Name, Surname, and Rank or Profession of Father. Name, and Maiden Surname of Mother.	If a regular Marriage, Signatures of officiating Minister and Witnesses. If irregular, Date of Conviction, Decree of Declarator, or Sheriff's Warrant.	When & W and Signat
	1903,	William Dundas Clark	28	15	William Clark	Warrant of	19
	on the Twenty-ninth	Aerated Water Bottler		Canal	Railway Engine Driver	Sheriff -	Jul
	day of July	(Bachelor)		Street,		- Substitute	
	at 14 Tay Street,			Perth.	Elizabeth Clark	of	
	Perth.				M.S. Moir	Perthshire	At Per
	By declaration in				(dec)	dated	
191	presence of Kate	Spinster Michelle Lennie	23	Almond Villa	David Lorimer	29th July 1903.	
	Smith, Tyer's Assistant	Domestic Servant		Glasgow	Marow		J.R
	4 South Street Perth.	(Spinster)		Road,	(dec)		Alex
	and Henry Clark			Perth.	Helen Lorimer		
	Butcher, 20 New Row.				M.S. Smith		
	Perth.				(dec)		
	1903,	(Signed)		82	Maurice Forde	(Signed)	19
	on the Thirty-first	J. Forde	28	South	Miner	James H. Cameron	Aug
	day of July	Tunnel Miner		Street,		Minister.	
	at 52 South Street,	(Bachelor)		Perth.	Mary Forde		
	Perth.				M.S. Conner		At Pe
					(dec)	(Signed)	
92	After Banns	(Signed)		82	Alexander Cumming	Tom Baker	
	According to the	Mary Croll	30	South	Chemist's Porter	Witness.	J.R
	forms of the			Street,		Ann Malloch	Ale
	United Free	(Widow)		Perth.	Mary Cumming	Witness.	
	Church of Scotland.				M.S. Gow		
					(dec)		

(1.) (2.) (3.) (4.) (5.) (6.)

M

James Bridges, Reg

Our Approach: Key Points

- Forming Conjugal Family Units (CFUs)
  - instead of linking individuals
- Using high-dimensional similarity search (M-Tree)
  - instead of blocking

Forming CFUs

- Conjugal Family Units (CFUs)
- Inspiration: Demographers
- Form families directly
  - Sibling groups
  - Identified by the names of the parents
  - and their date & place of marriage

Forming CFUs

Birth	Marriage	Death
Father's Forename	Groom's Forename	Father's Forename
Father's Surname	Groom's Surname	Father's Surname
Mother's Forename	Bride's Forename	Mother's Forename
Mother's Surname	Bride's Surname	Mother's Surname
Parent's Date of Marriage	Date of Marriage	
Parent's Place of Marriage	Place of Marriage	

Our Approach: Steps

- Forming sibling groups (Birth – Birth)
  - Link people whose parents are the same
- Associating the marriage (Birth – Marriage)
  - Construct phantom marriage records for each birth
  - Find the closest real marriage record
- Linking death events into familial groups
  - Harder since date and place of marriage are missing
- Inter-generational linkage
  - Marriage – Birth (using phantom Birth records)
  - Marriage – Death (using phantom Death records)
  - Marriage – Marriage (using phantom Marriage records, using groom's parents and bride's parents)

Algorithm

- Construct Birth M-Tree
- For each birth record
  - Find the nearest neighbour (M-Tree)
  - Distance < Family forming threshold
    - Same family
- Construct Family M-Tree
- For each family
  - Find the nearest neighbour (M-Tree)
  - Distance < Family merging threshold
    - Merge families
- Hard limit on maximum family size

Algorithm

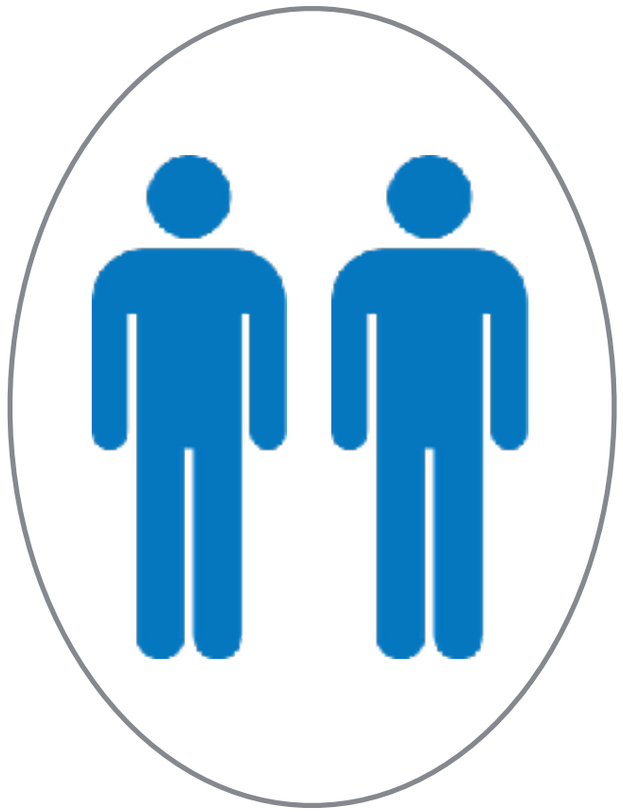
- Construct Birth M-Tree
- For each birth record
  - Find the nearest neighbour (M-Tree)
  - Distance < Family forming threshold
    - Same family
- Construct Family M-Tree
- For each family
  - Find the nearest neighbour (M-Tree)
  - Distance < Family merging threshold
    - Merge families
- Hard limit on maximum family size

Algorithm

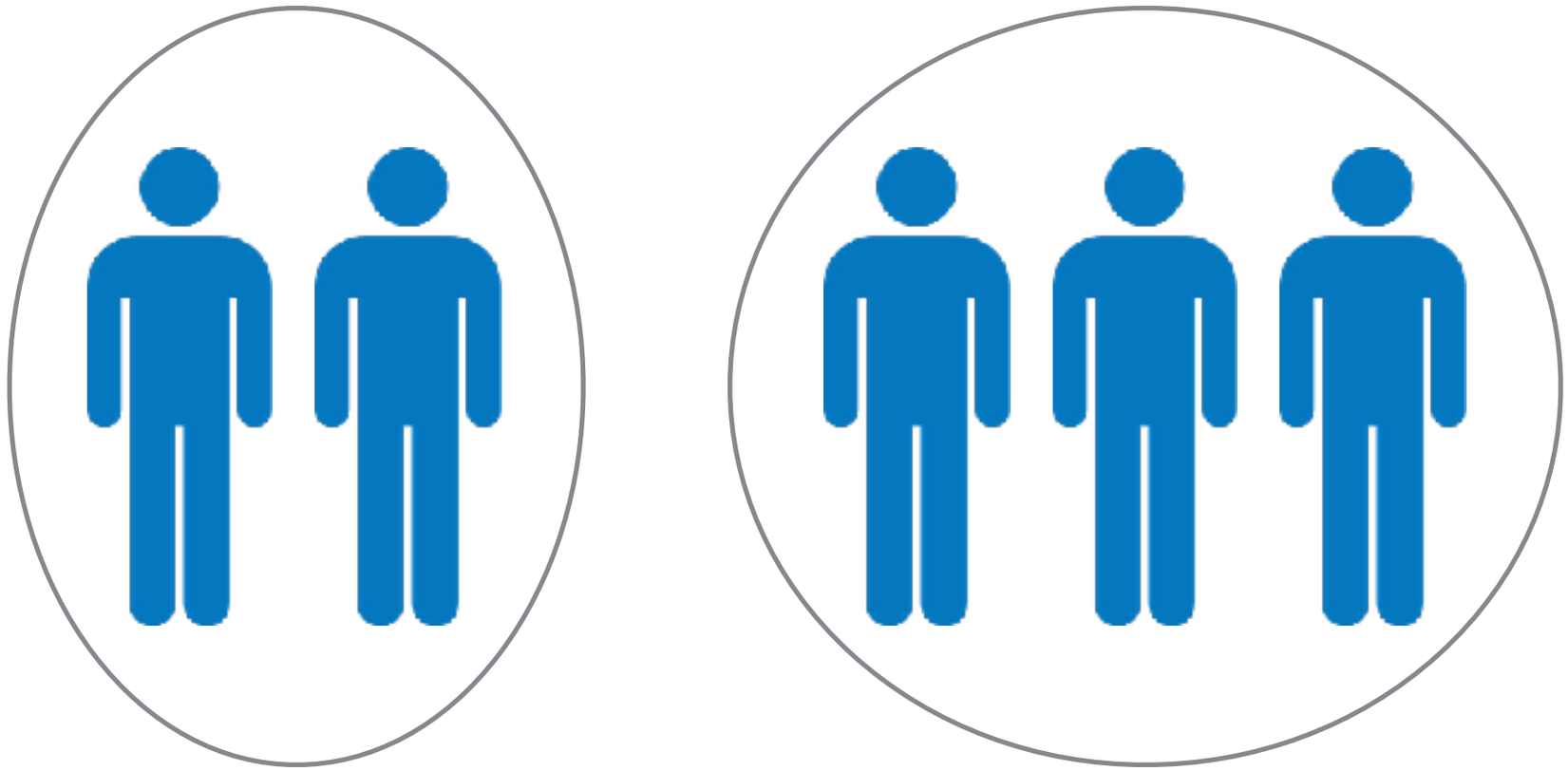
- Construct Birth M-Tree
- For each birth record
  - Find the nearest neighbour (M-Tree)
  - Distance < Family forming threshold
    - Same family
- Construct Family M-Tree
- For each family
  - Find the nearest neighbour (M-Tree)
  - Distance < Family merging threshold
    - Merge families
- Hard limit on maximum family size

Algorithm

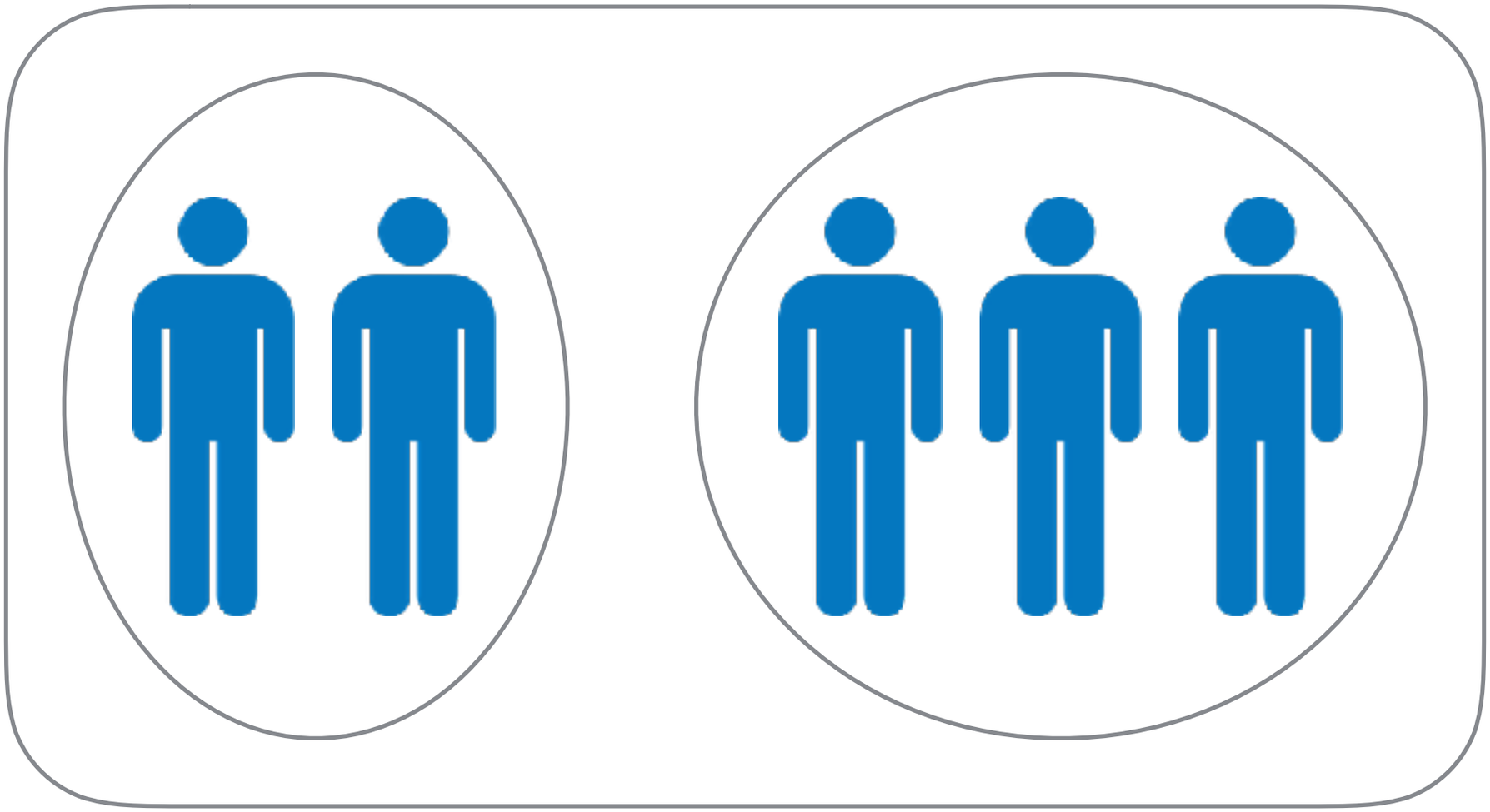
Algorithm



Algorithm



Algorithm



M-Tree

M-Tree

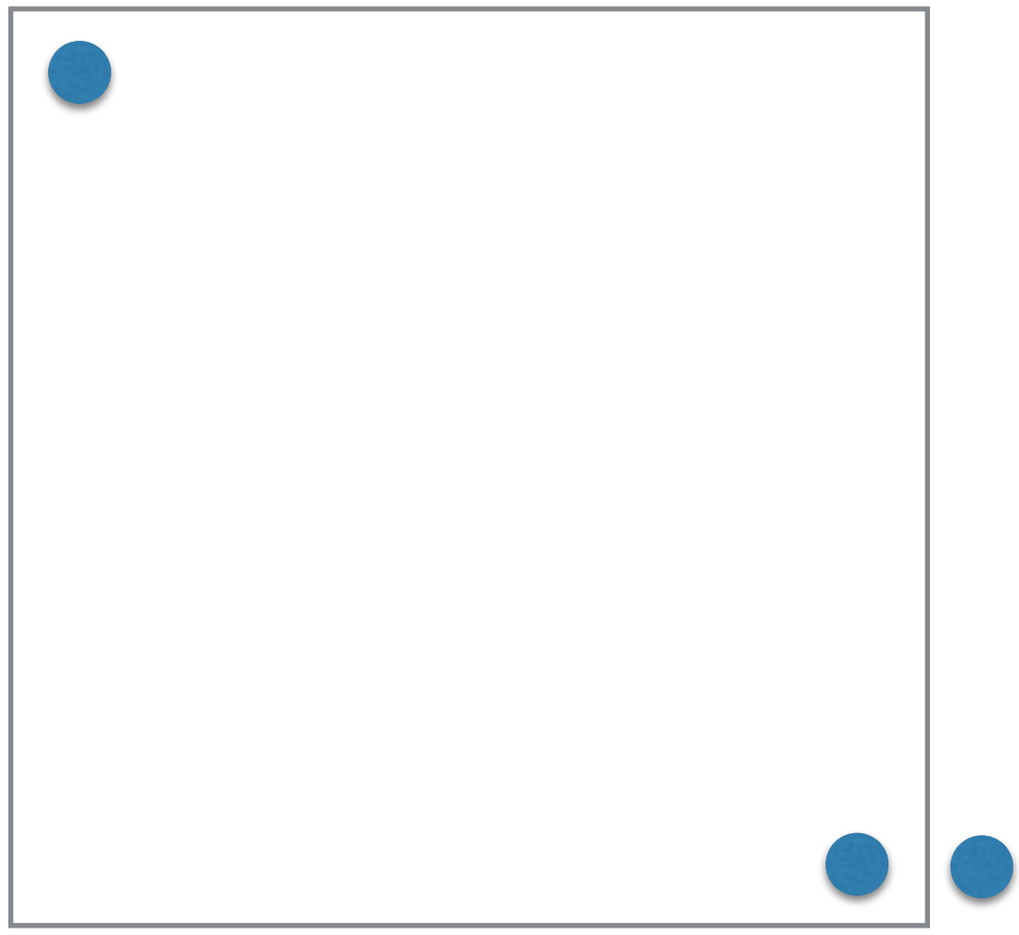
- A data structure which places similar objects close to each other
- Allows looking up neighbour objects efficiently

M-Tree

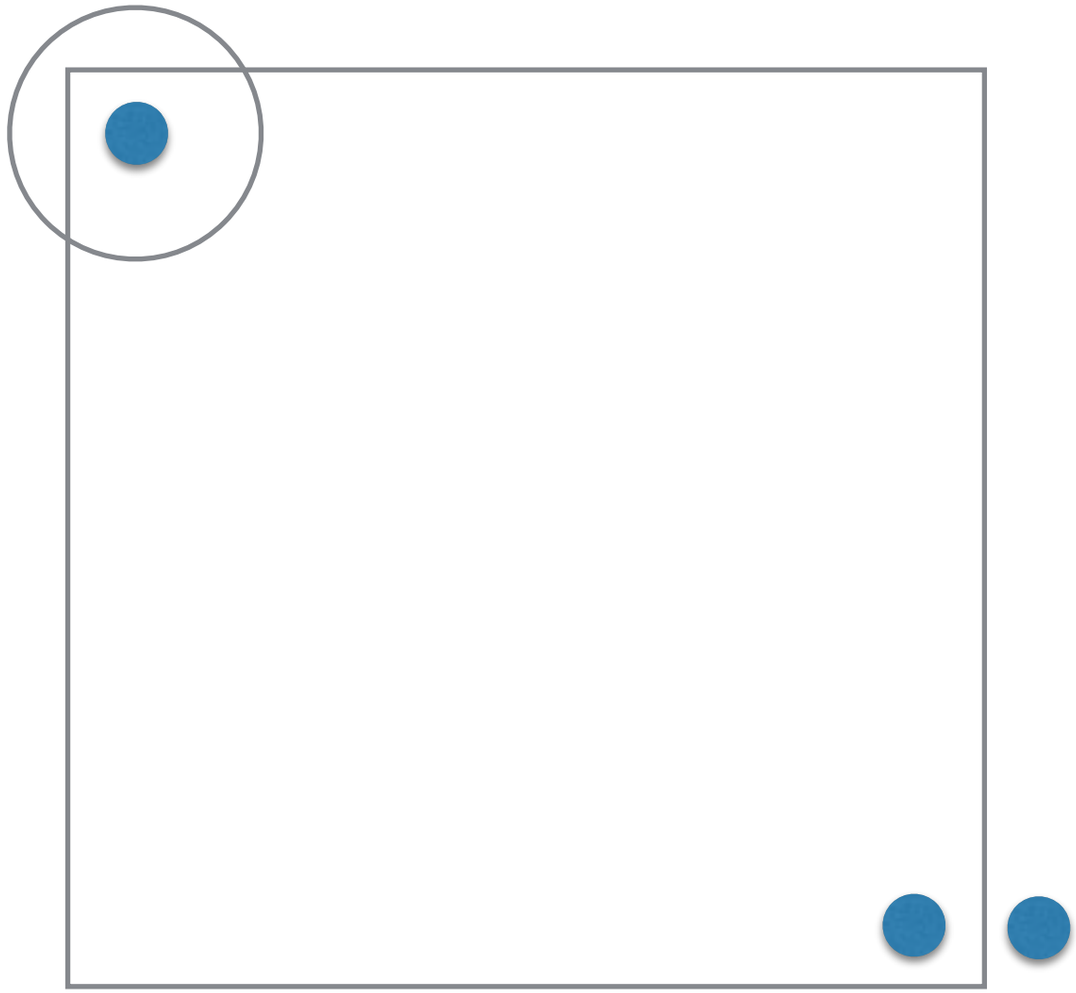
- A data structure which places similar objects close to each other
- Allows looking up neighbour objects efficiently

- Ciaccia, P., Patella, M. and Zezula, P., 1997.
  M-tree: An Effcient Access Method for Similarity Search in Metric Spaces.
  In Proceedings of the 23rd VLDB conference

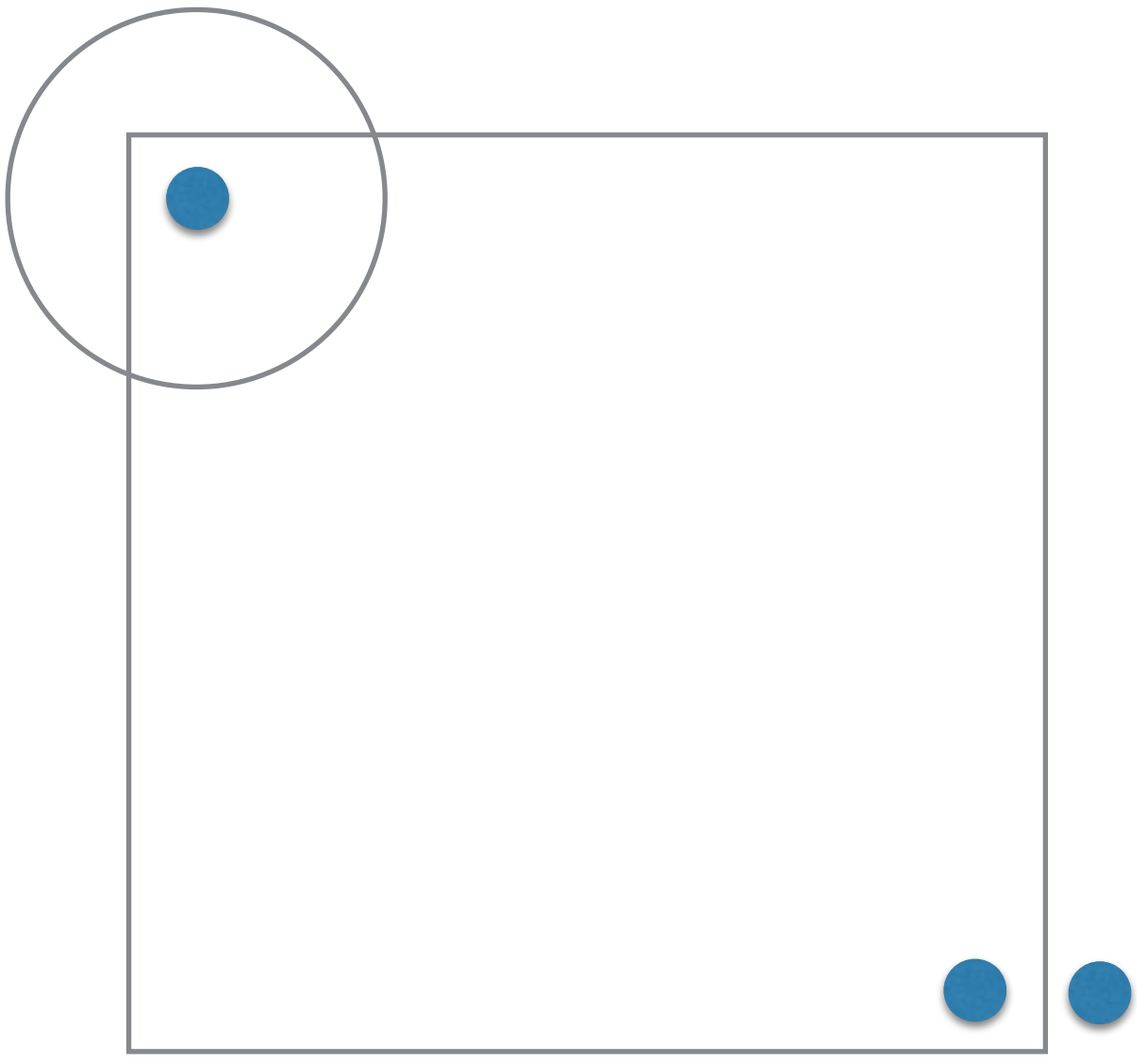
M-Tree vs Blocking



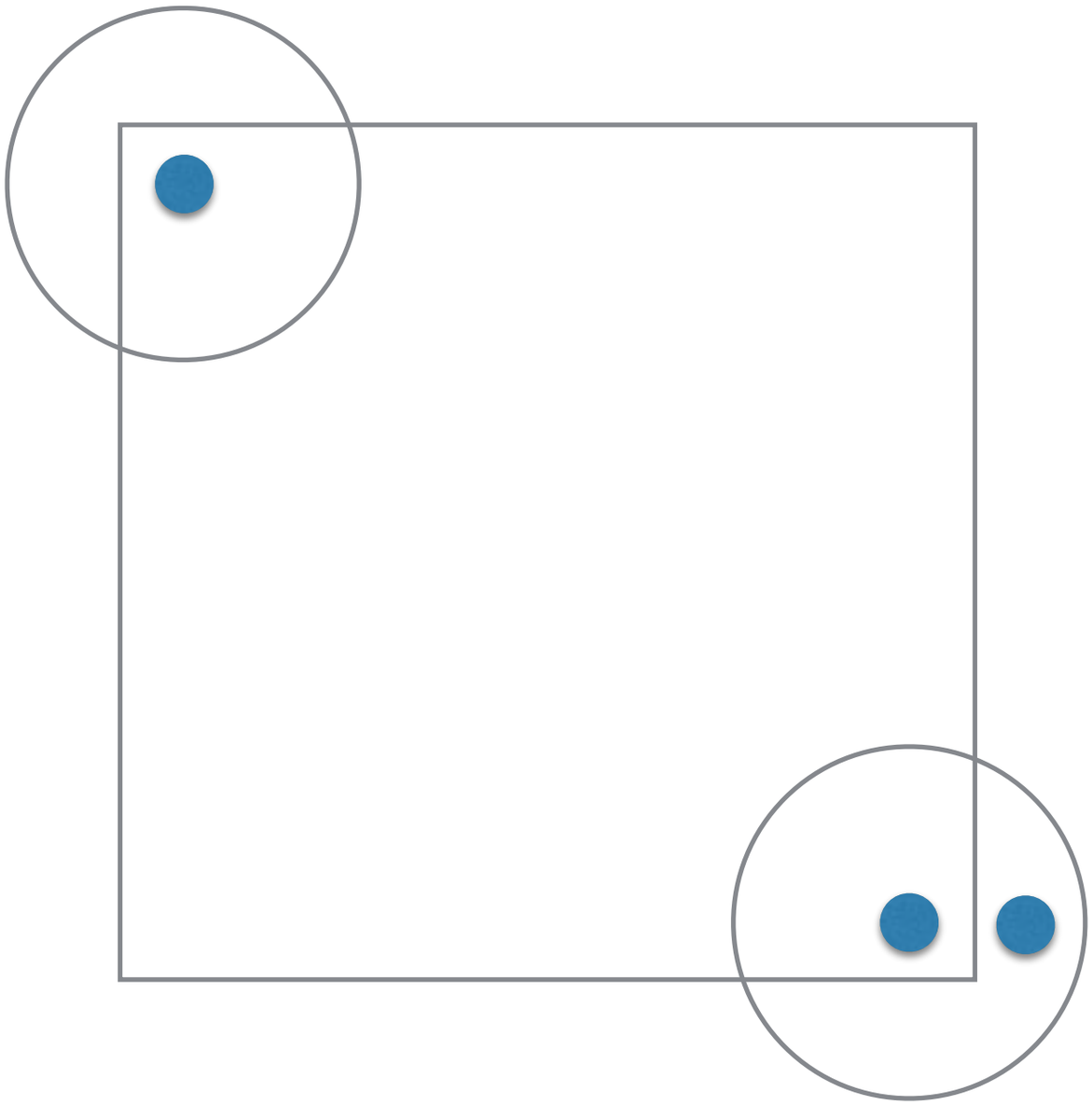
M-Tree vs Blocking



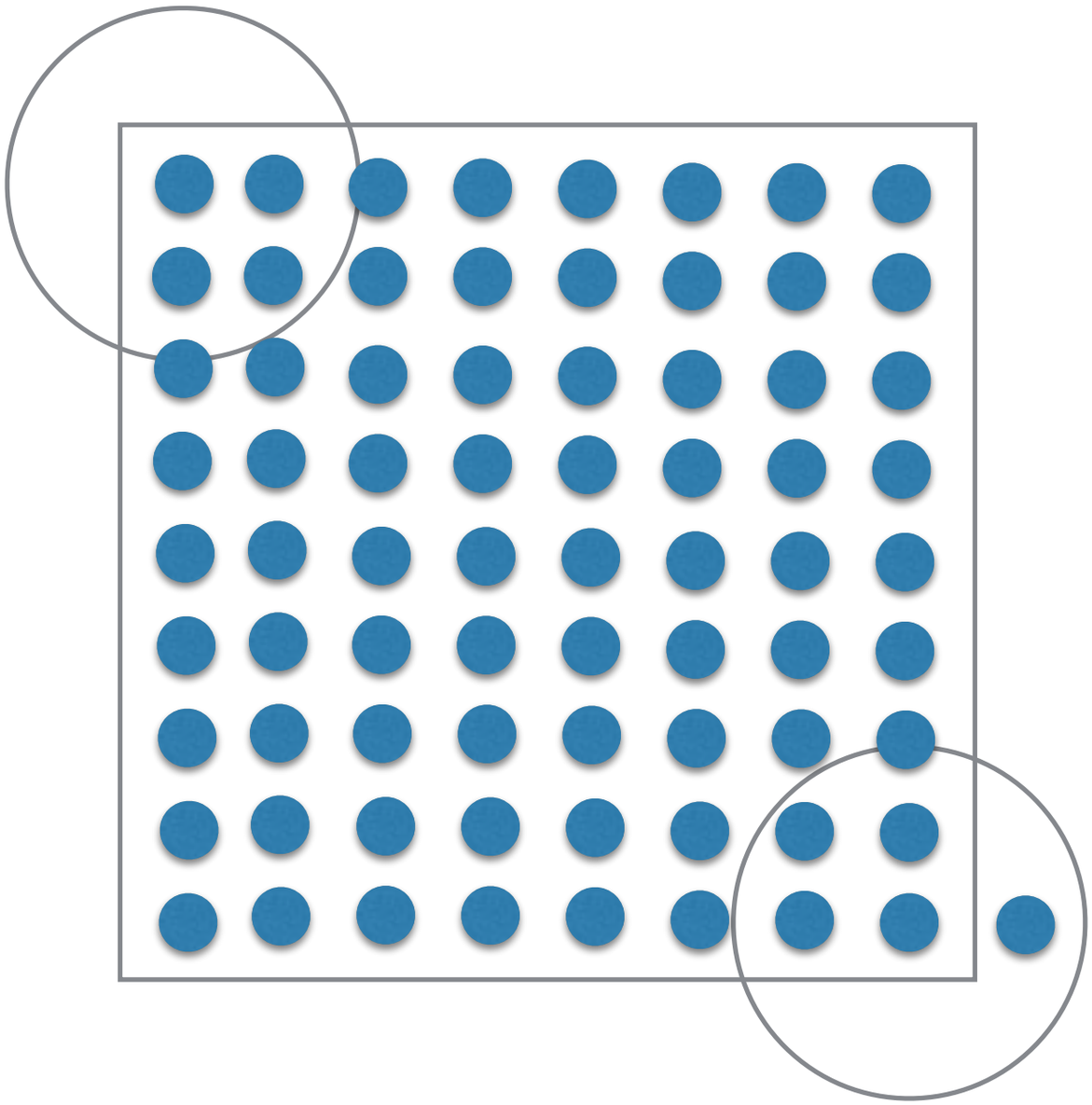
M-Tree vs Blocking

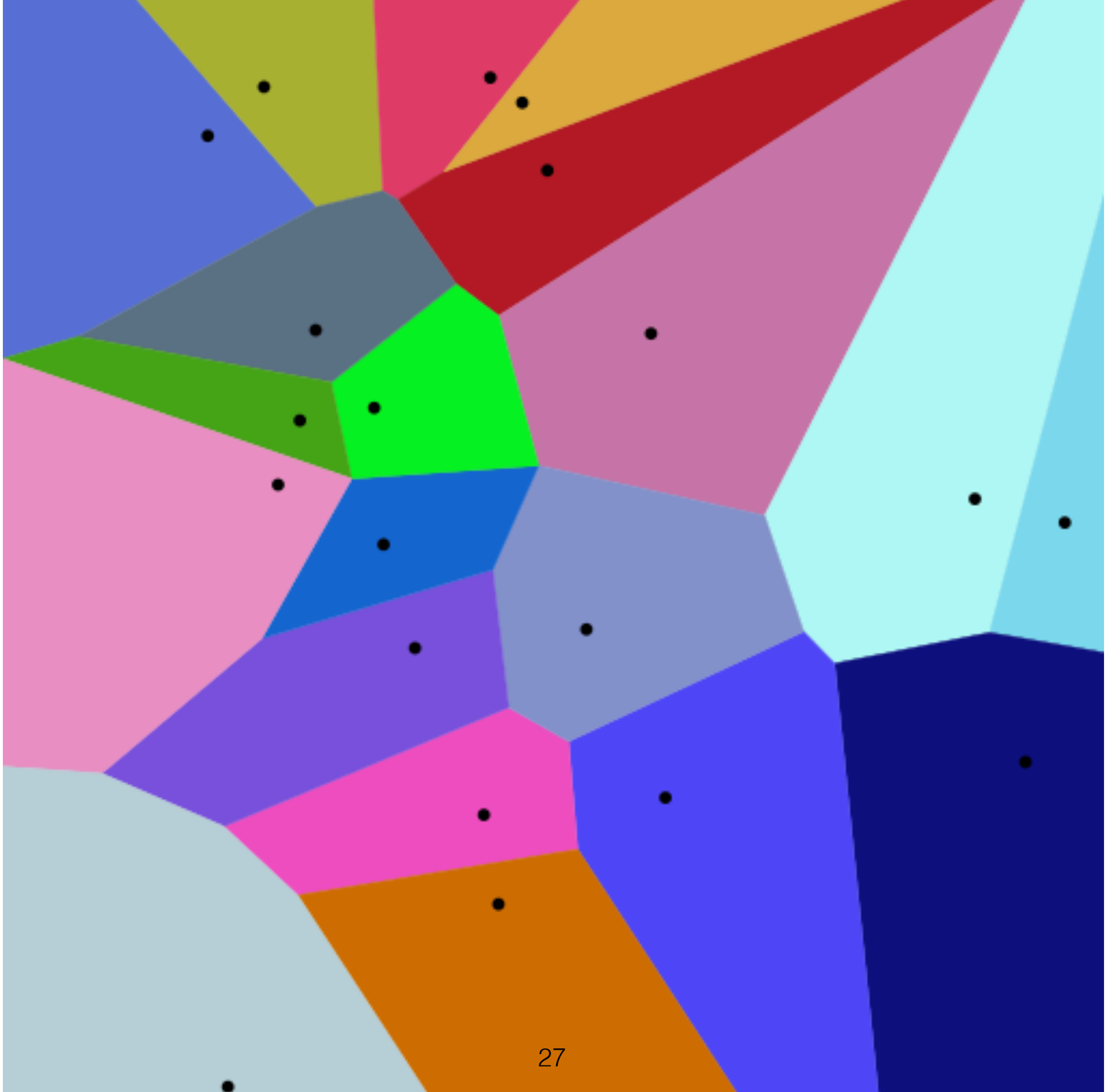


M-Tree vs Blocking



M-Tree vs Blocking





Exploratory Results

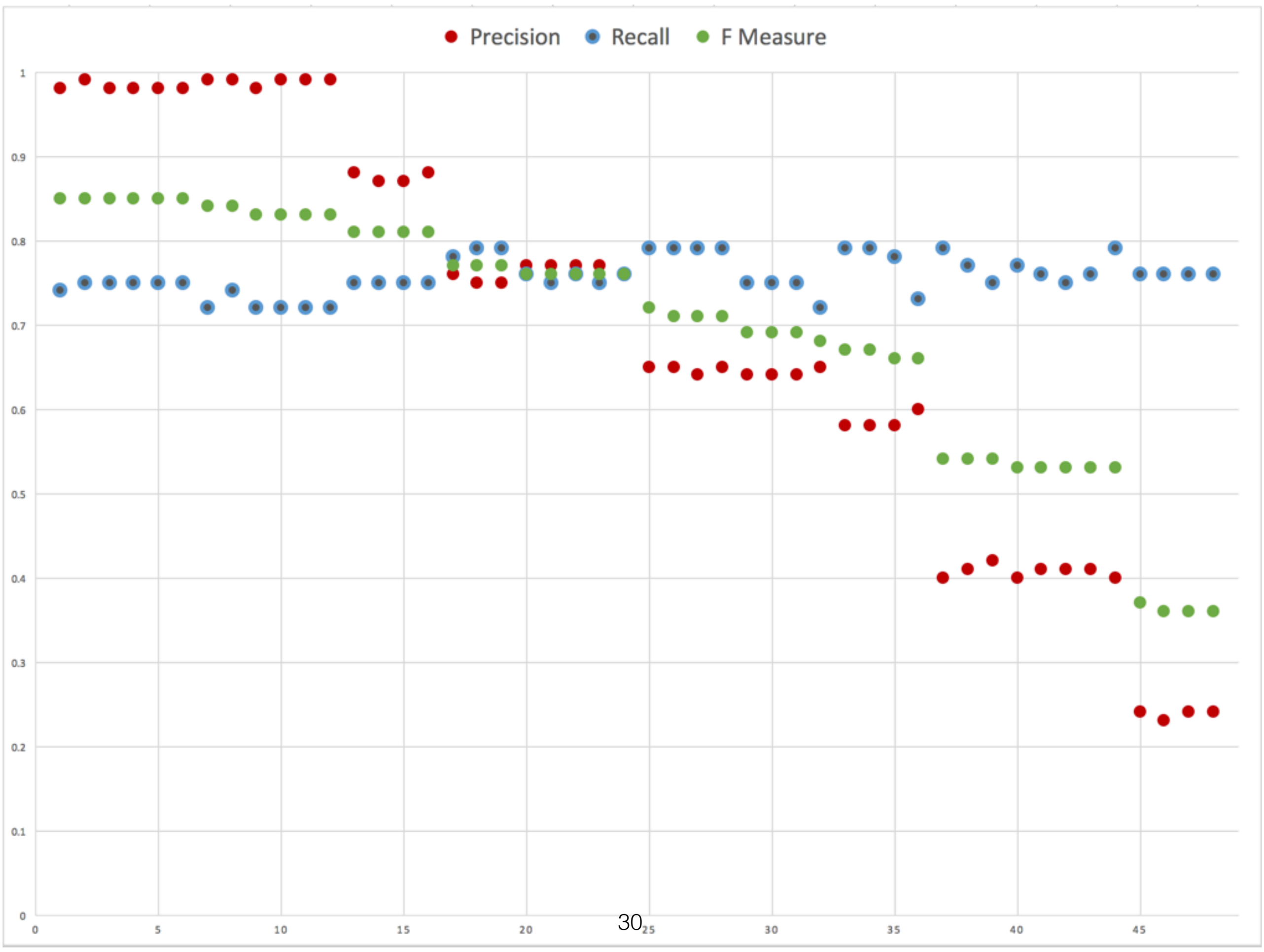
- We used Levenshtein as the distance metric
- We explored
  - Key fields
    - FFN-FLN-MFN-MLN
    - FFN-FLN-MFN-MLN-DOM-POM
- Original/Clean fields
- Thresholds
  - Family forming distance (3, 7, 9, 16)
  - Maximum family size (5, 10, 20)
  - Family merge distance (3, 7, 9, 16)

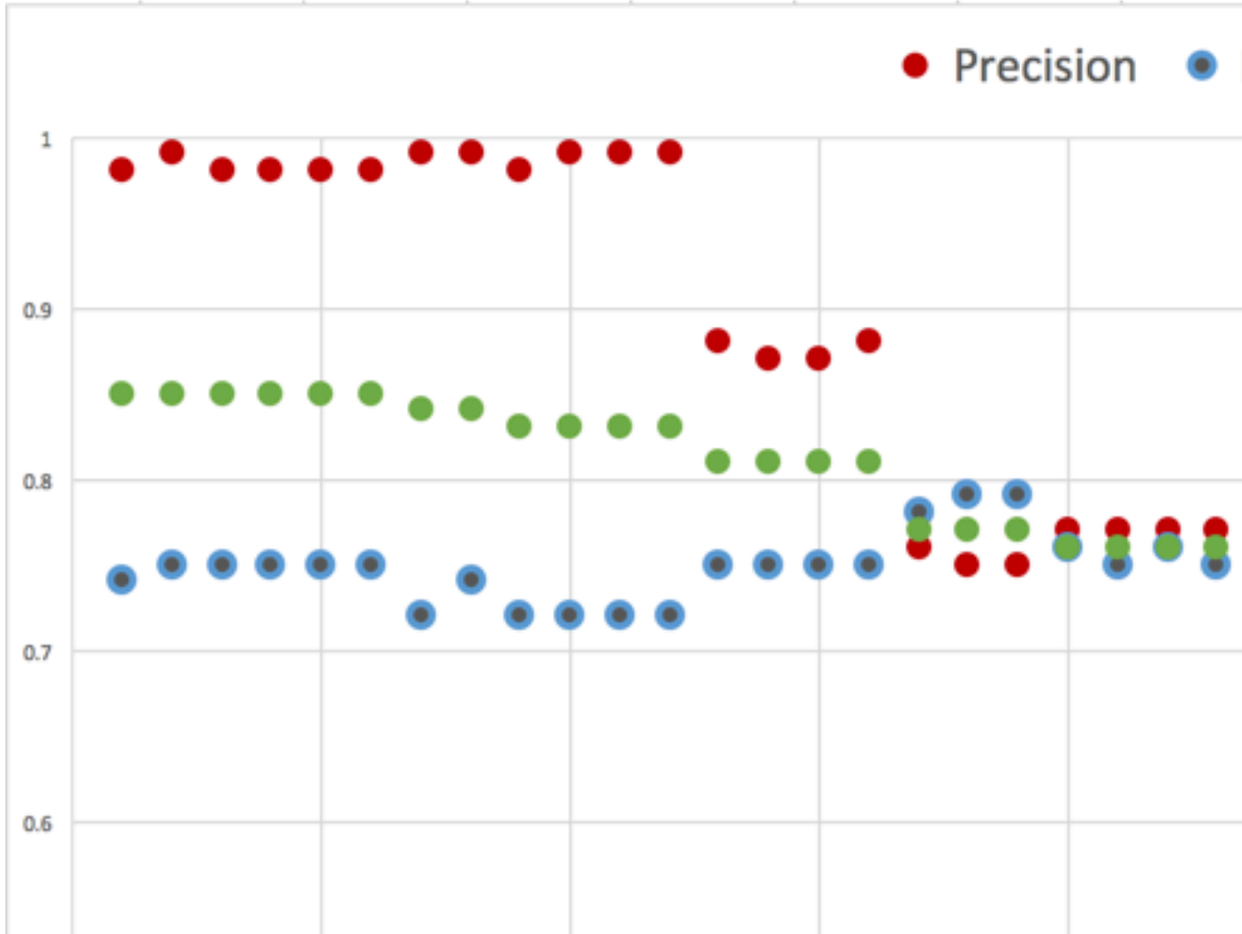
Linkage Quality

- Precision = TP / (TP + FP)
  - When we claim a link, how often are we right?
  - Perfect precision: find 1 correct link

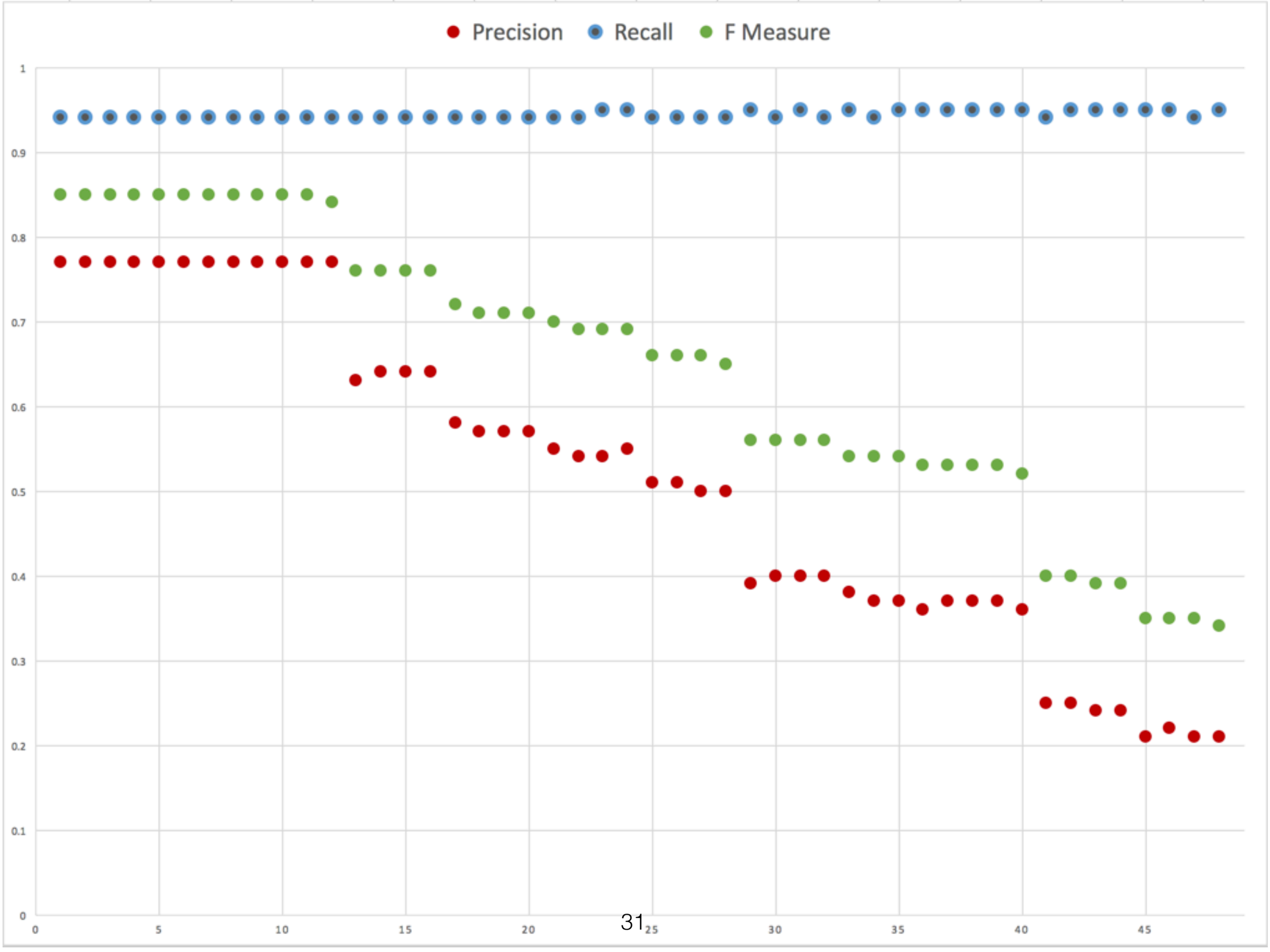
- Recall = TP / (TP + FN)
  - How complete are we?
  - Perfect recall: claim everything is a link!

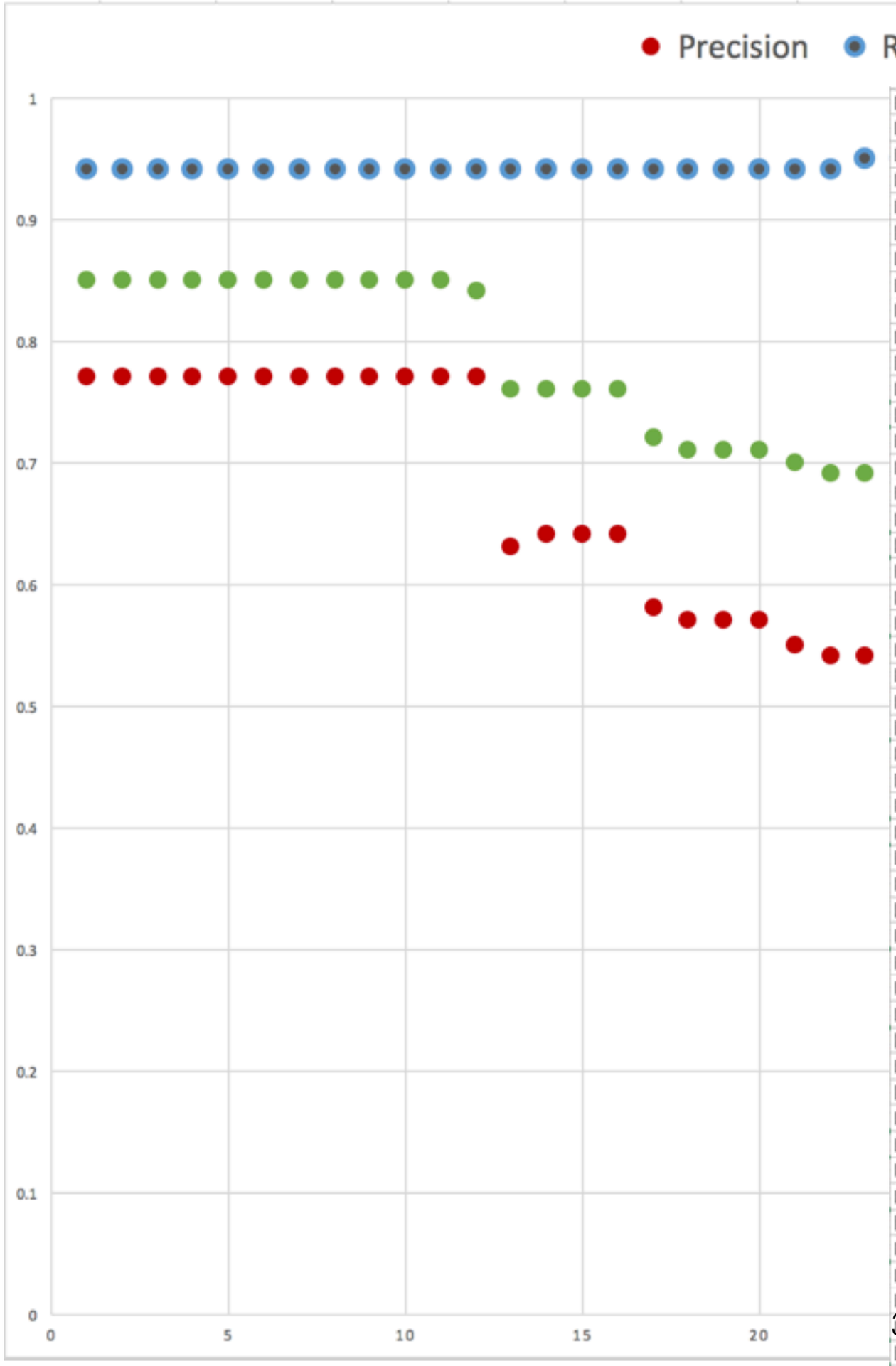
- F Measure = 2 * Precision * Recall
  / (Precision + Recall)


Precision
Recall
F Measure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25
30
35
40
45

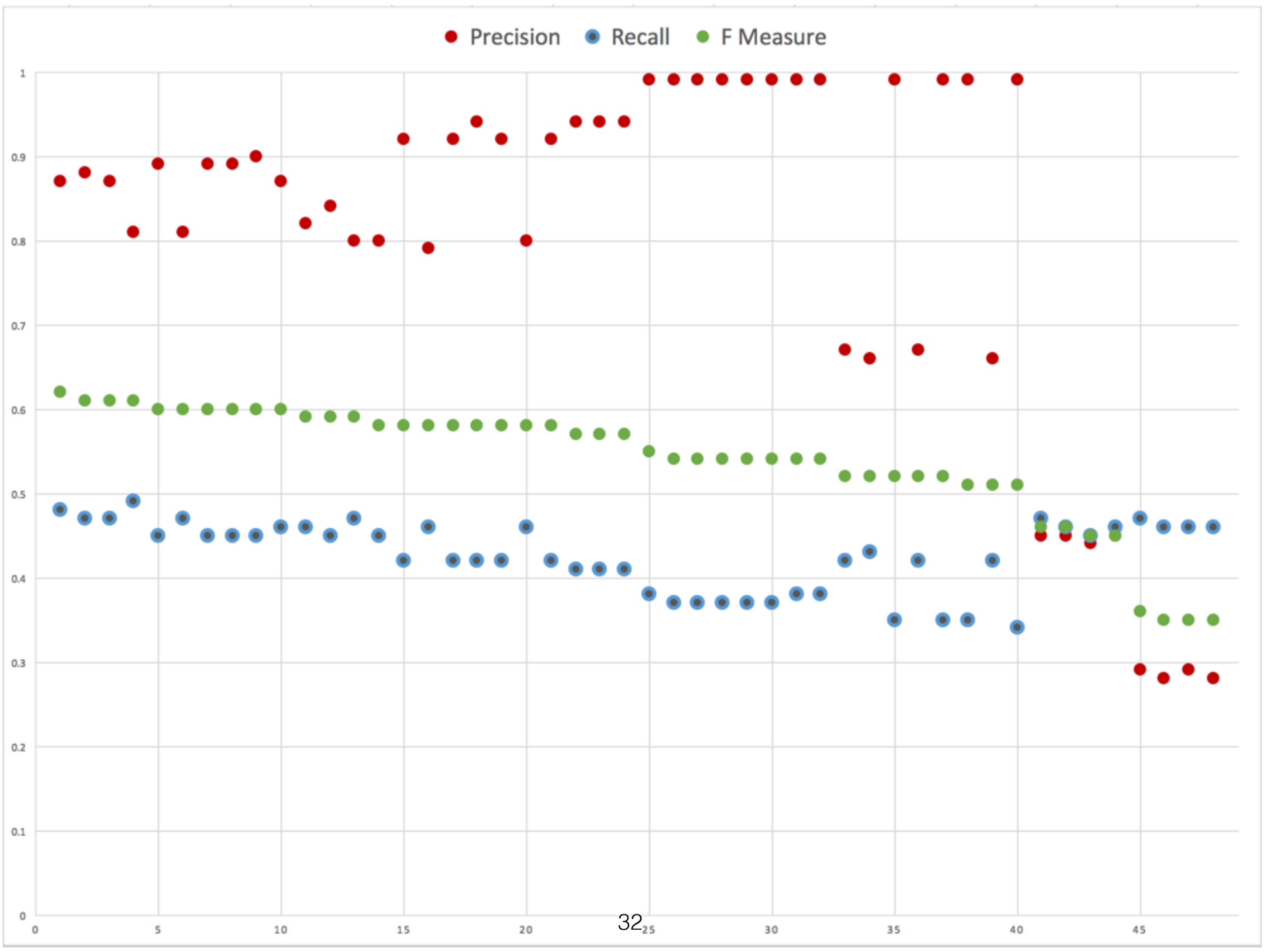


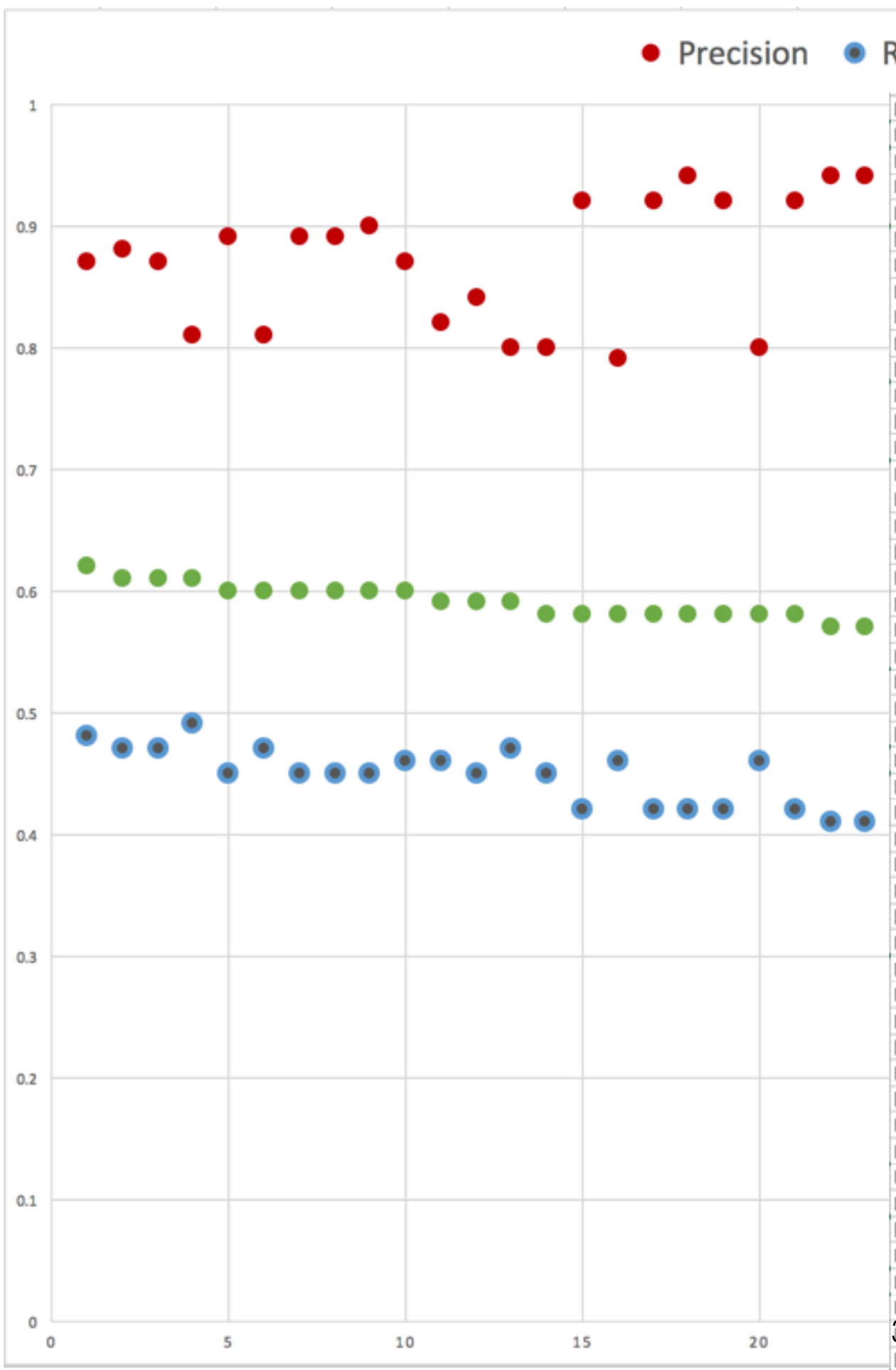
DataSet	Keys	Distance	Family Si	Merge Di	Precision	Recall	F Measur
kilmarnock	NamesOriginal	16	20	3	0.98	0.74	0.85
kilmarnock	NamesOriginal	3	10	3	0.99	0.75	0.85
kilmarnock	NamesOriginal	3	20	3	0.98	0.75	0.85
kilmarnock	NamesOriginal	7	20	3	0.98	0.75	0.85
kilmarnock	NamesOriginal	9	10	3	0.98	0.75	0.85
kilmarnock	NamesOriginal	9	20	3	0.98	0.75	0.85
kilmarnock	NamesOriginal	16	5	3	0.99	0.72	0.84
kilmarnock	NamesOriginal	7	10	3	0.99	0.74	0.84
kilmarnock	NamesOriginal	16	10	3	0.98	0.72	0.83
kilmarnock	NamesOriginal	3	5	3	0.99	0.72	0.83
kilmarnock	NamesOriginal	7	5	3	0.99	0.72	0.83
kilmarnock	NamesOriginal	9	5	3	0.99	0.72	0.83
kilmarnock	NamesOriginal	16	5	7	0.88	0.75	0.81
kilmarnock	NamesOriginal	3	5	7	0.87	0.75	0.81
kilmarnock	NamesOriginal	7	5	7	0.87	0.75	0.81
kilmarnock	NamesOriginal	9	5	7	0.88	0.75	0.81
kilmarnock	NamesOriginal	16	10	7	0.76	0.78	0.77
kilmarnock	NamesOriginal	3	10	7	0.75	0.79	0.77
kilmarnock	NamesOriginal	7	10	7	0.75	0.79	0.77
kilmarnock	NamesOriginal	16	5	9	0.77	0.75	0.76
kilmarnock	NamesOriginal	3	5	9	0.77	0.75	0.76
kilmarnock	NamesOriginal	7	5	9	0.77	0.75	0.76
kilmarnock	NamesOriginal	9	10	7	0.77	0.75	0.76
kilmarnock	NamesOriginal	9	5	9	0.76	0.75	0.76
kilmarnock	NamesOriginal	7	20	7	0.65	0.79	0.72
kilmarnock	NamesOriginal	16	20	7	0.65	0.79	0.71
kilmarnock	NamesOriginal	3	20	7	0.64	0.79	0.71
kilmarnock	NamesOriginal	9	20	7	0.65	0.79	0.71
kilmarnock	NamesOriginal	16	5	16	0.64	0.75	0.69
kilmarnock	NamesOriginal	3	5	16	0.64	0.75	0.69
kilmarnock	NamesOriginal	7	5	16	0.64	0.75	0.69
kilmarnock	NamesOriginal	9	5	16	0.65	0.72	0.68
kilmarnock	NamesOriginal	16	10	9	0.58	0.79	0.67
kilmarnock	NamesOriginal	7	10	9	0.58	0.79	0.67
kilmarnock	NamesOriginal	3	10	9	0.58	0.78	0.66
kilmarnock	NamesOriginal	9	10	9	0.6	0.73	0.66
kilmarnock	NamesOriginal	16	20	9	0.4	0.79	0.54
kilmarnock	NamesOriginal	3	10	16	0.41	0.77	0.54
kilmarnock	NamesOriginal	3	20	9	0.42	0.75	0.54
kilmarnock	NamesOriginal	16	10	16	0.4	0.77	0.53
kilmarnock	NamesOriginal	7	10	16	0.41	0.75	0.53
kilmarnock	NamesOriginal	7	20	9	0.41	0.75	0.53
kilmarnock	NamesOriginal	9	10	16	0.41	0.75	0.53
kilmarnock	NamesOriginal	9	20	9	0.4	0.79	0.53
kilmarnock	NamesOriginal	9	20	16	0.24	0.75	0.37
kilmarnock	NamesOriginal	16	20	16	0.23	0.75	0.36
kilmarnock	NamesOriginal	3	20	16	0.24	0.75	0.36
kilmarnock	NamesOriginal	7	20	16	0.24	0.75	0.36


Precision
Recall
F Measure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25
30
35
40
45

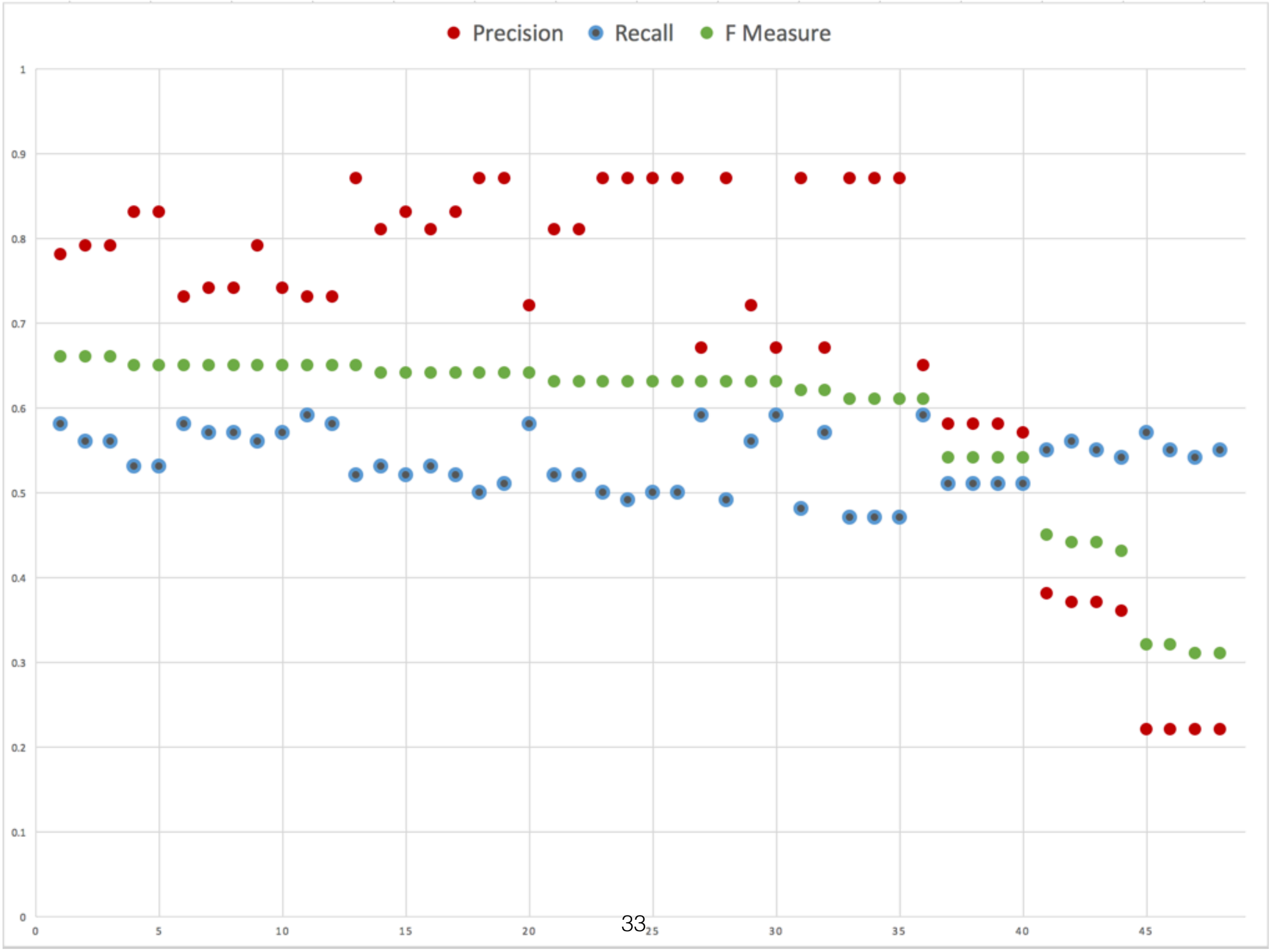


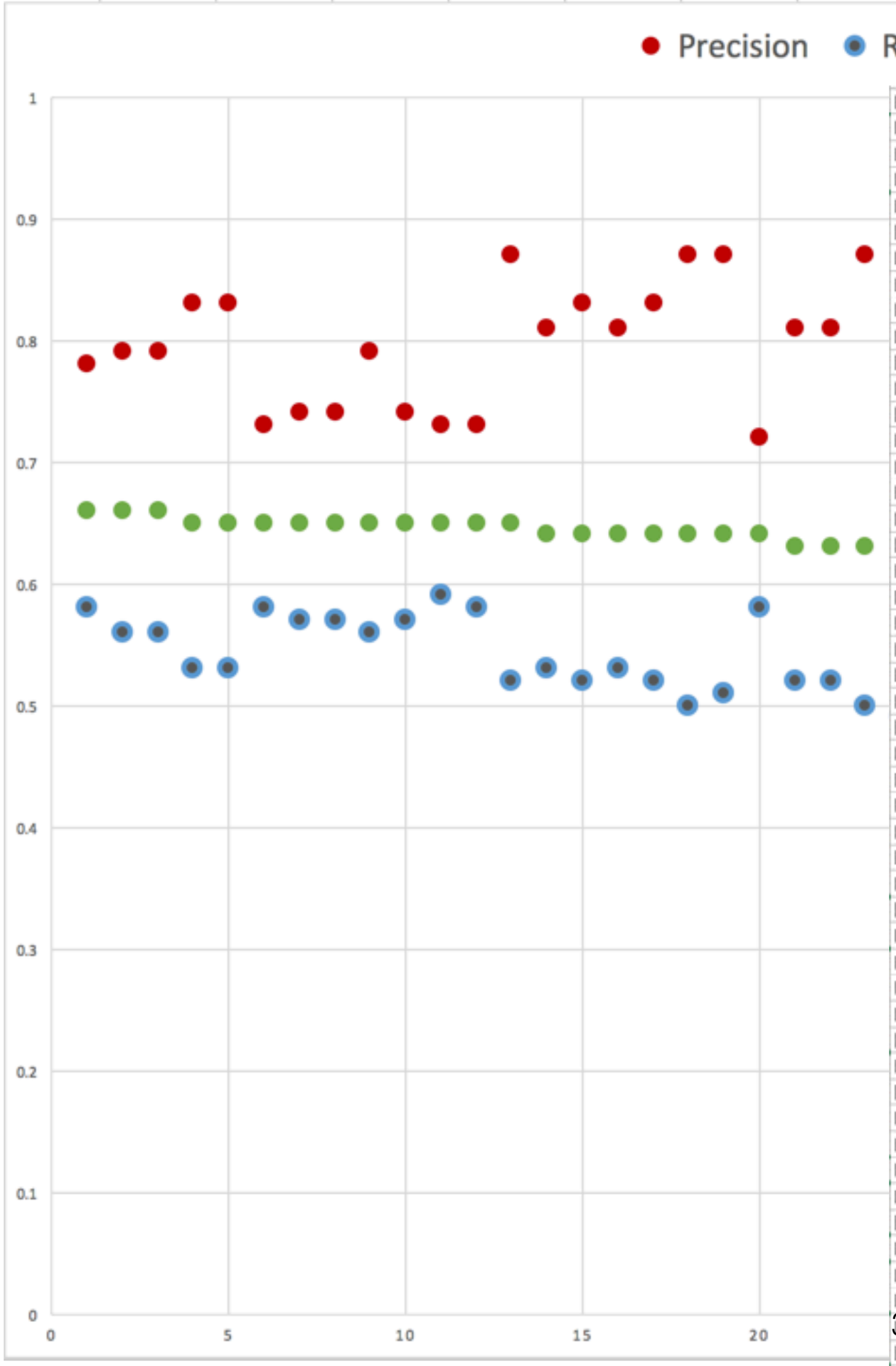
DataSet	Keys	Distance	Family Si	Merge Di	Precision	Recall	F Measur
kilmarnock	NamesClean	16	10	3	0.77	0.94	0.85
kilmarnock	NamesClean	16	20	3	0.77	0.94	0.85
kilmarnock	NamesClean	16	5	3	0.77	0.94	0.85
kilmarnock	NamesClean	3	10	3	0.77	0.94	0.85
kilmarnock	NamesClean	3	20	3	0.77	0.94	0.85
kilmarnock	NamesClean	3	5	3	0.77	0.94	0.85
kilmarnock	NamesClean	7	10	3	0.77	0.94	0.85
kilmarnock	NamesClean	7	5	3	0.77	0.94	0.85
kilmarnock	NamesClean	9	10	3	0.77	0.94	0.85
kilmarnock	NamesClean	9	20	3	0.77	0.94	0.85
kilmarnock	NamesClean	9	5	3	0.77	0.94	0.85
kilmarnock	NamesClean	7	20	3	0.77	0.94	0.84
kilmarnock	NamesClean	16	5	7	0.63	0.94	0.76
kilmarnock	NamesClean	3	5	7	0.64	0.94	0.76
kilmarnock	NamesClean	7	5	7	0.64	0.94	0.76
kilmarnock	NamesClean	9	5	7	0.64	0.94	0.76
kilmarnock	NamesClean	9	5	9	0.58	0.94	0.72
kilmarnock	NamesClean	16	5	9	0.57	0.94	0.71
kilmarnock	NamesClean	3	5	9	0.57	0.94	0.71
kilmarnock	NamesClean	7	5	9	0.57	0.94	0.71
kilmarnock	NamesClean	9	5	16	0.55	0.94	0.7
kilmarnock	NamesClean	16	5	16	0.54	0.94	0.69
kilmarnock	NamesClean	3	5	16	0.54	0.95	0.69
kilmarnock	NamesClean	7	5	16	0.55	0.95	0.69
kilmarnock	NamesClean	16	10	7	0.51	0.94	0.66
kilmarnock	NamesClean	3	10	7	0.51	0.94	0.66
kilmarnock	NamesClean	9	10	7	0.5	0.94	0.66
kilmarnock	NamesClean	7	10	7	0.5	0.94	0.65
kilmarnock	NamesClean	16	10	9	0.39	0.95	0.56
kilmarnock	NamesClean	3	10	9	0.4	0.94	0.56
kilmarnock	NamesClean	7	10	9	0.4	0.95	0.56
kilmarnock	NamesClean	9	10	9	0.4	0.94	0.56
kilmarnock	NamesClean	16	20	7	0.38	0.95	0.54
kilmarnock	NamesClean	7	20	7	0.37	0.94	0.54
kilmarnock	NamesClean	9	20	7	0.37	0.95	0.54
kilmarnock	NamesClean	16	10	16	0.36	0.95	0.53
kilmarnock	NamesClean	3	10	16	0.37	0.95	0.53
kilmarnock	NamesClean	3	20	7	0.37	0.95	0.53
kilmarnock	NamesClean	7	10	16	0.37	0.95	0.53
kilmarnock	NamesClean	9	10	16	0.36	0.95	0.52
kilmarnock	NamesClean	7	20	9	0.25	0.94	0.4
kilmarnock	NamesClean	9	20	9	0.25	0.95	0.4
kilmarnock	NamesClean	16	20	9	0.24	0.95	0.39
kilmarnock	NamesClean	3	20	9	0.24	0.95	0.39
kilmarnock	NamesClean	16	20	16	0.21	0.95	0.35
kilmarnock	NamesClean	3	20	16	0.22	0.95	0.35
kilmarnock	NamesClean	9	20	16	0.21	0.94	0.35
kilmarnock	NamesClean	7	20	16	0.21	0.95	0.34


Precision
Recall
F Measure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25
30
35
40
45

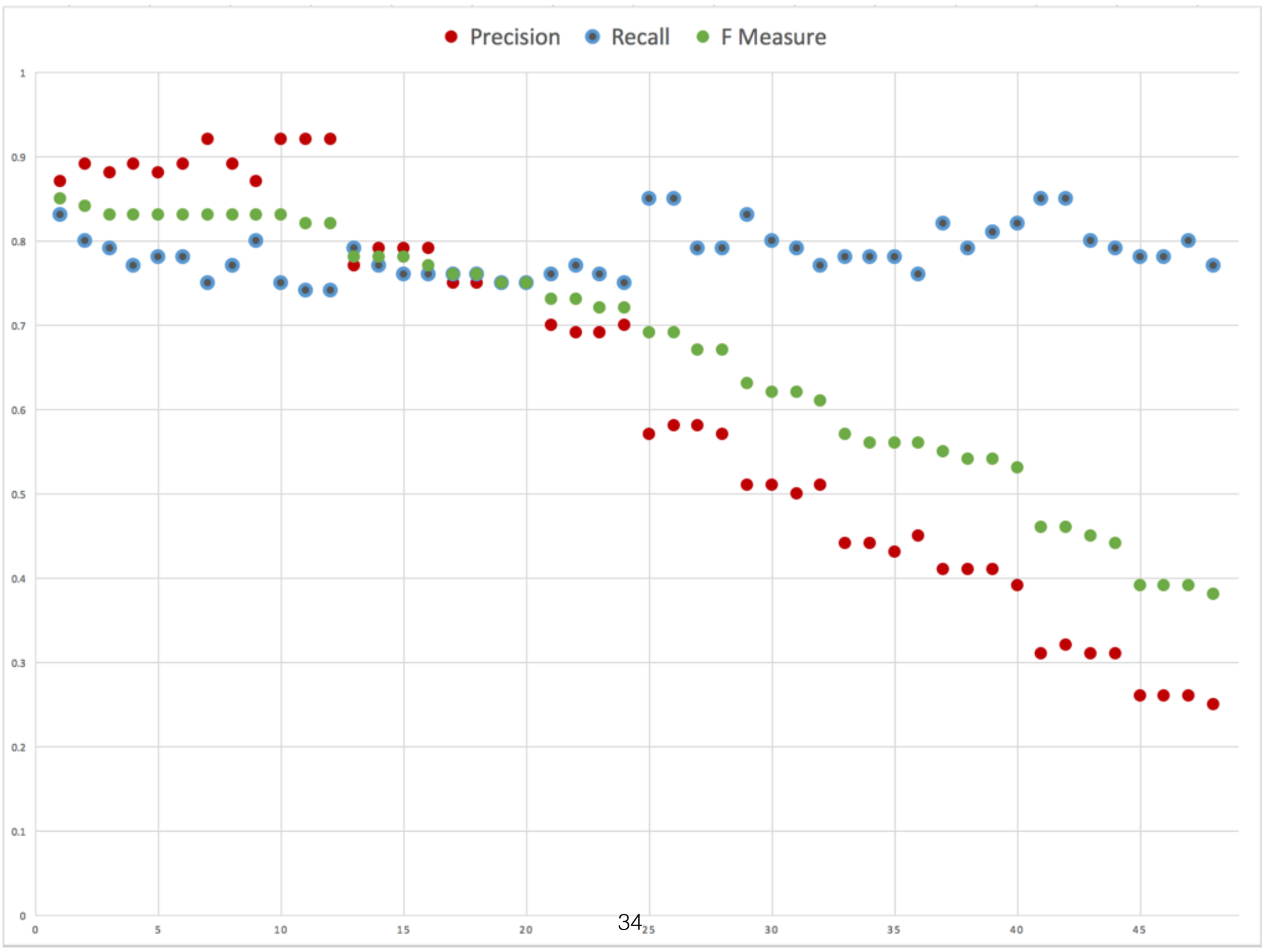


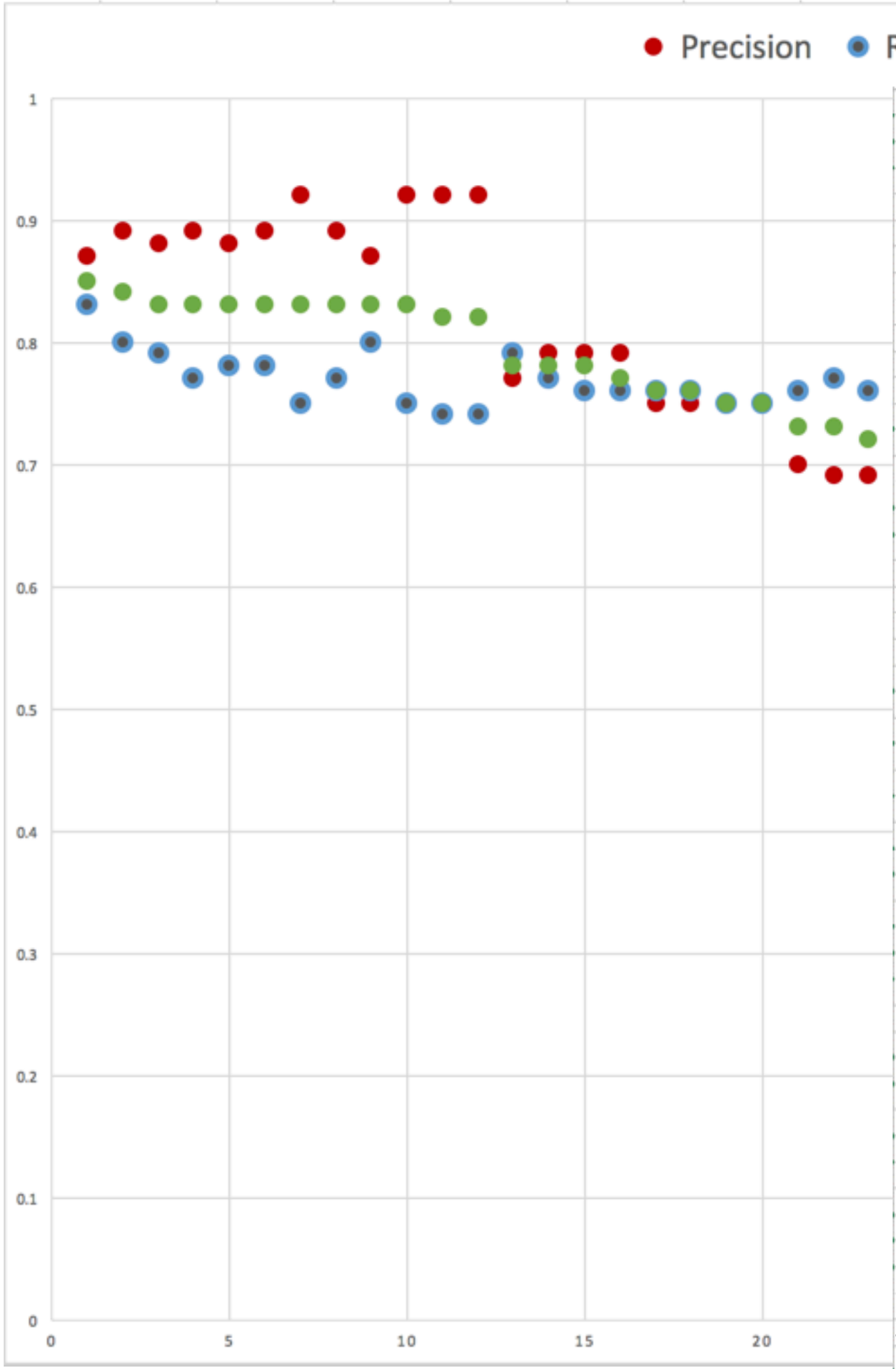
DataSet	Keys	Distance	Family Si	Merge Di	Precision	Recall	F Measur
kilmarnock	NamesMarriageOrig nal	3	10	9	0.87	0.48	0.62
kilmarnock	NamesMarriageOrig nal	16	10	9	0.88	0.47	0.61
kilmarnock	NamesMarriageOrig nal	7	10	9	0.87	0.47	0.61
kilmarnock	NamesMarriageOrig nal	9	20	9	0.81	0.49	0.61
kilmarnock	NamesMarriageOrig nal	16	10	7	0.89	0.45	0.6
kilmarnock	NamesMarriageOrig nal	16	20	9	0.81	0.47	0.6
kilmarnock	NamesMarriageOrig nal	3	10	7	0.89	0.45	0.6
kilmarnock	NamesMarriageOrig nal	7	10	7	0.89	0.45	0.6
kilmarnock	NamesMarriageOrig nal	9	10	7	0.9	0.45	0.6
kilmarnock	NamesMarriageOrig nal	9	10	9	0.87	0.46	0.6
kilmarnock	NamesMarriageOrig nal	3	20	7	0.82	0.46	0.59
kilmarnock	NamesMarriageOrig nal	7	20	7	0.84	0.45	0.59
kilmarnock	NamesMarriageOrig nal	7	20	9	0.8	0.47	0.59
kilmarnock	NamesMarriageOrig nal	16	20	7	0.8	0.45	0.58
kilmarnock	NamesMarriageOrig nal	16	5	9	0.92	0.42	0.58
kilmarnock	NamesMarriageOrig nal	3	20	9	0.79	0.46	0.58
kilmarnock	NamesMarriageOrig nal	3	5	9	0.92	0.42	0.58
kilmarnock	NamesMarriageOrig nal	7	5	7	0.94	0.42	0.58
kilmarnock	NamesMarriageOrig nal	7	5	9	0.92	0.42	0.58
kilmarnock	NamesMarriageOrig nal	9	20	7	0.8	0.46	0.58
kilmarnock	NamesMarriageOrig nal	9	5	9	0.92	0.42	0.58
kilmarnock	NamesMarriageOrig nal	16	5	7	0.94	0.41	0.57
kilmarnock	NamesMarriageOrig nal	3	5	7	0.94	0.41	0.57
kilmarnock	NamesMarriageOrig nal	9	5	7	0.94	0.41	0.57
kilmarnock	NamesMarriageOrig nal	3	20	3	0.99	0.38	0.55
kilmarnock	NamesMarriageOrig nal	16	10	3	0.99	0.37	0.54
kilmarnock	NamesMarriageOrig nal	16	20	3	0.99	0.37	0.54
kilmarnock	NamesMarriageOrig nal	3	10	3	0.99	0.37	0.54
kilmarnock	NamesMarriageOrig nal	7	10	3	0.99	0.37	0.54
kilmarnock	NamesMarriageOrig nal	7	20	3	0.99	0.37	0.54
kilmarnock	NamesMarriageOrig nal	9	10	3	0.99	0.38	0.54
kilmarnock	NamesMarriageOrig nal	9	20	3	0.99	0.38	0.54
kilmarnock	NamesMarriageOrig nal	16	5	16	0.67	0.42	0.52
kilmarnock	NamesMarriageOrig nal	3	5	16	0.66	0.43	0.52
kilmarnock	NamesMarriageOrig nal	3	5	3	0.99	0.35	0.52
kilmarnock	NamesMarriageOrig nal	7	5	16	0.67	0.42	0.52
kilmarnock	NamesMarriageOrig nal	7	5	3	0.99	0.35	0.52
kilmarnock	NamesMarriageOrig nal	16	5	3	0.99	0.35	0.51
kilmarnock	NamesMarriageOrig nal	9	5	16	0.66	0.42	0.51
kilmarnock	NamesMarriageOrig nal	9	5	3	0.99	0.34	0.51
kilmarnock	NamesMarriageOrig nal	16	10	16	0.45	0.47	0.46
kilmarnock	NamesMarriageOrig nal	7	10	16	0.45	0.46	0.46
kilmarnock	NamesMarriageOrig nal	3	10	16	0.44	0.45	0.45
kilmarnock	NamesMarriageOrig nal	9	10	16	0.45	0.46	0.45
kilmarnock	NamesMarriageOrig nal	16	20	16	0.29	0.47	0.36
kilmarnock	NamesMarriageOrig nal	3	20	16	0.28	0.46	0.35
kilmarnock	NamesMarriageOrig nal	7	20	16	0.29	0.46	0.35
kilmarnock	NamesMarriageOrig nal	9	20	16	0.28	0.46	0.35


Precision
Recall
F Measure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25
30
35
40
45

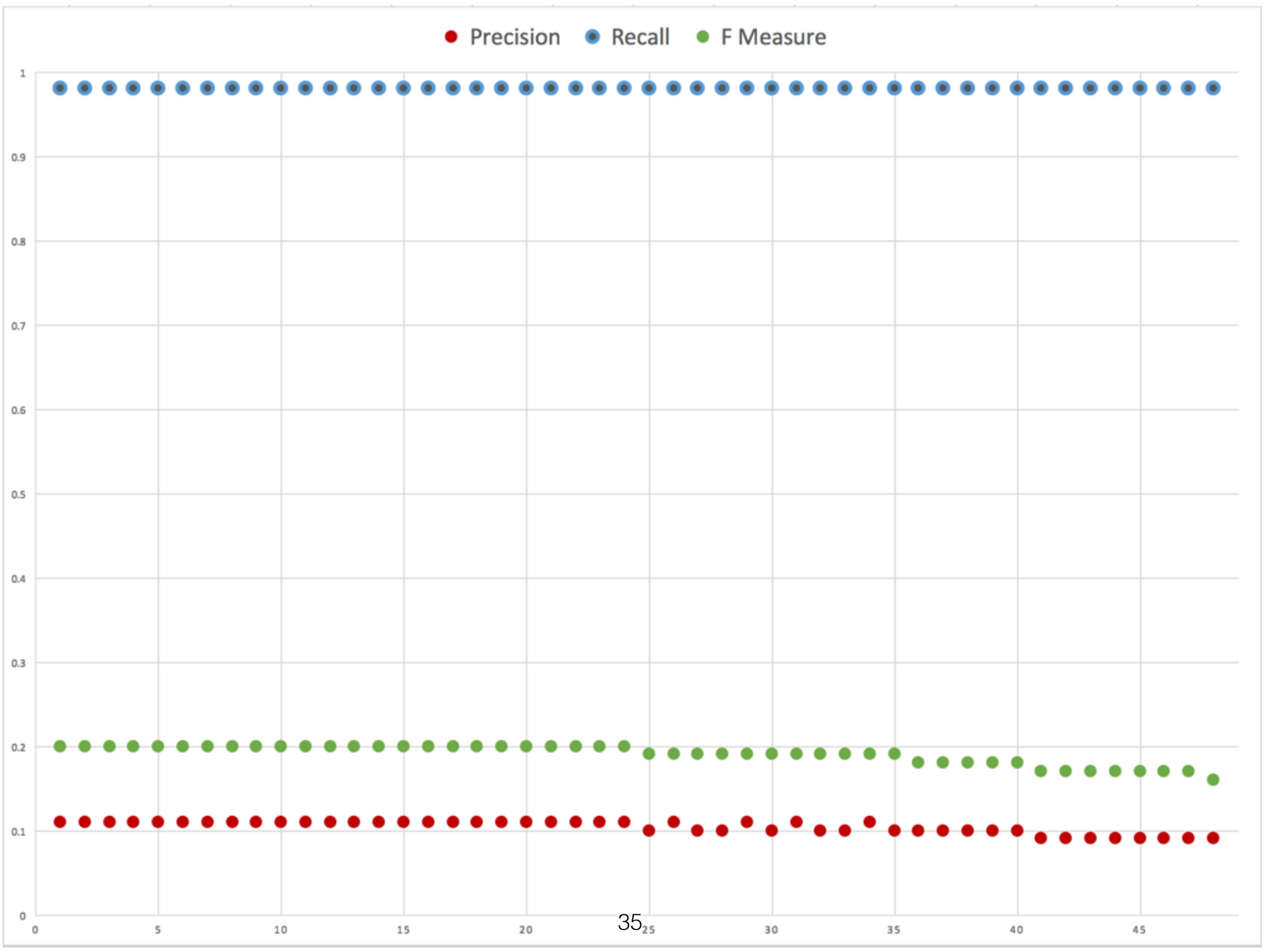


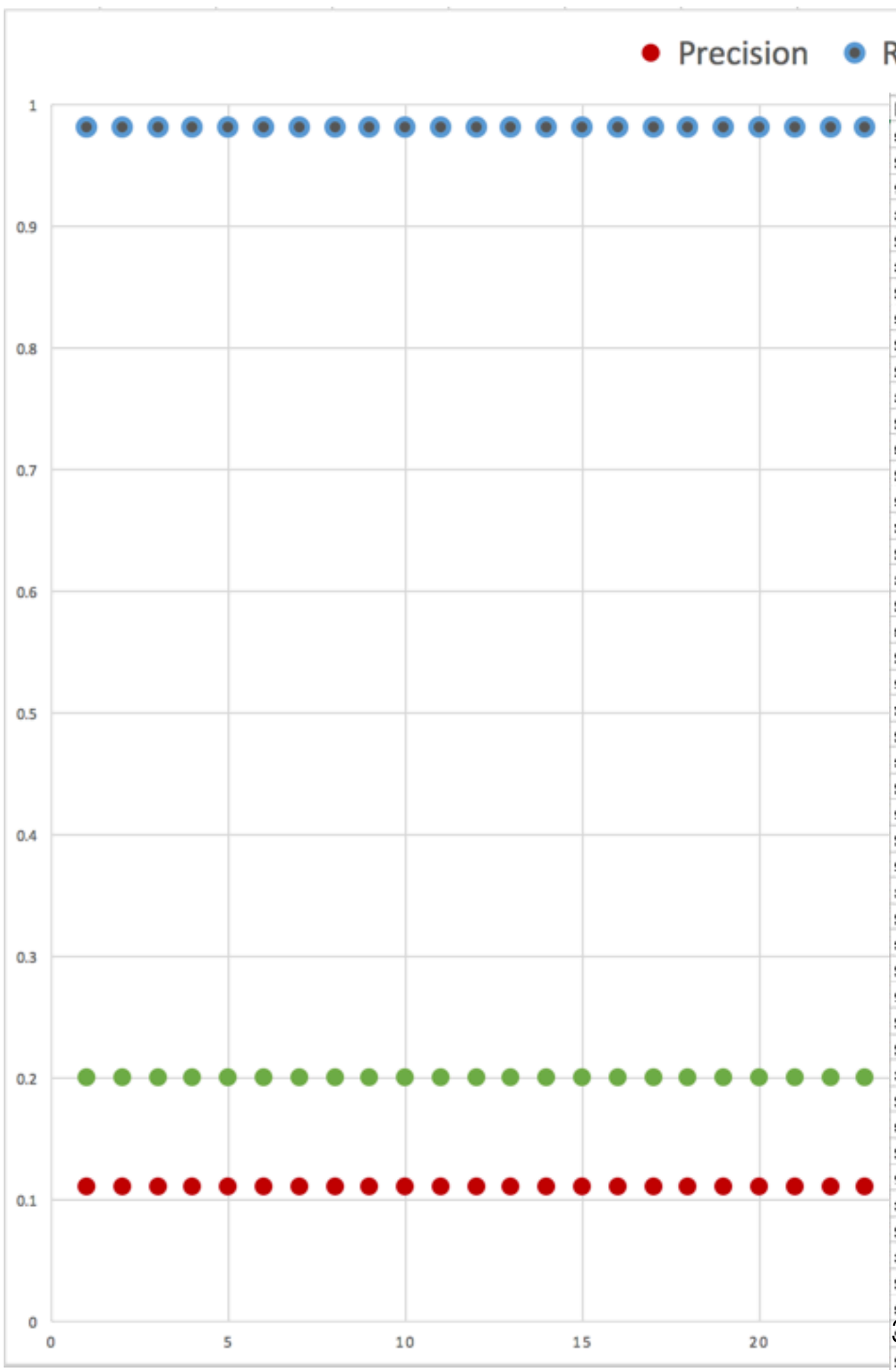
DataSet	Keys	Distance	Family Si	Merge Di	Precision	Recall	F Measur
kilmarnock	NamesMarriageClean	3	10	7	0.78	0.58	0.66
kilmarnock	NamesMarriageClean	7	10	7	0.79	0.56	0.66
kilmarnock	NamesMarriageClean	9	10	7	0.79	0.56	0.66
kilmarnock	NamesMarriageClean	3	5	7	0.83	0.53	0.65
kilmarnock	NamesMarriageClean	9	5	7	0.83	0.53	0.65
kilmarnock	NamesMarriageClean	3	10	9	0.73	0.58	0.65
kilmarnock	NamesMarriageClean	7	10	9	0.74	0.57	0.65
kilmarnock	NamesMarriageClean	9	10	9	0.74	0.57	0.65
kilmarnock	NamesMarriageClean	16	10	7	0.79	0.56	0.65
kilmarnock	NamesMarriageClean	16	10	9	0.74	0.57	0.65
kilmarnock	NamesMarriageClean	3	20	7	0.73	0.59	0.65
kilmarnock	NamesMarriageClean	7	20	7	0.73	0.58	0.65
kilmarnock	NamesMarriageClean	9	20	3	0.87	0.52	0.65
kilmarnock	NamesMarriageClean	3	5	9	0.81	0.53	0.64
kilmarnock	NamesMarriageClean	7	5	7	0.83	0.52	0.64
kilmarnock	NamesMarriageClean	9	5	9	0.81	0.53	0.64
kilmarnock	NamesMarriageClean	16	5	7	0.83	0.52	0.64
kilmarnock	NamesMarriageClean	7	10	3	0.87	0.5	0.64
kilmarnock	NamesMarriageClean	16	20	3	0.87	0.51	0.64
kilmarnock	NamesMarriageClean	16	20	7	0.72	0.58	0.64
kilmarnock	NamesMarriageClean	7	5	9	0.81	0.52	0.63
kilmarnock	NamesMarriageClean	16	5	9	0.81	0.52	0.63
kilmarnock	NamesMarriageClean	3	10	3	0.87	0.5	0.63
kilmarnock	NamesMarriageClean	9	10	3	0.87	0.49	0.63
kilmarnock	NamesMarriageClean	16	10	3	0.87	0.5	0.63
kilmarnock	NamesMarriageClean	3	20	3	0.87	0.5	0.63
kilmarnock	NamesMarriageClean	3	20	9	0.67	0.59	0.63
kilmarnock	NamesMarriageClean	7	20	3	0.87	0.49	0.63
kilmarnock	NamesMarriageClean	9	20	7	0.72	0.56	0.63
kilmarnock	NamesMarriageClean	9	20	9	0.67	0.59	0.63
kilmarnock	NamesMarriageClean	7	5	3	0.87	0.48	0.62
kilmarnock	NamesMarriageClean	7	20	9	0.67	0.57	0.62
kilmarnock	NamesMarriageClean	3	5	3	0.87	0.47	0.61
kilmarnock	NamesMarriageClean	9	5	3	0.87	0.47	0.61
kilmarnock	NamesMarriageClean	16	5	3	0.87	0.47	0.61
kilmarnock	NamesMarriageClean	16	20	9	0.65	0.59	0.61
kilmarnock	NamesMarriageClean	3	5	16	0.58	0.51	0.54
kilmarnock	NamesMarriageClean	7	5	16	0.58	0.51	0.54
kilmarnock	NamesMarriageClean	9	5	16	0.58	0.51	0.54
kilmarnock	NamesMarriageClean	16	5	16	0.57	0.51	0.54
kilmarnock	NamesMarriageClean	7	10	16	0.38	0.55	0.45
kilmarnock	NamesMarriageClean	3	10	16	0.37	0.56	0.44
kilmarnock	NamesMarriageClean	9	10	16	0.37	0.55	0.44
kilmarnock	NamesMarriageClean	16	10	16	0.36	0.54	0.43
kilmarnock	NamesMarriageClean	3	20	16	0.22	0.57	0.32
kilmarnock	NamesMarriageClean	16	20	16	0.22	0.55	0.32
kilmarnock	NamesMarriageClean	7	20	16	0.22	0.54	0.31
kilmarnock	NamesMarriageClean	9	20	16	0.22	0.55	0.31


Precision
Recall
F Measure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25
30
35
40
45

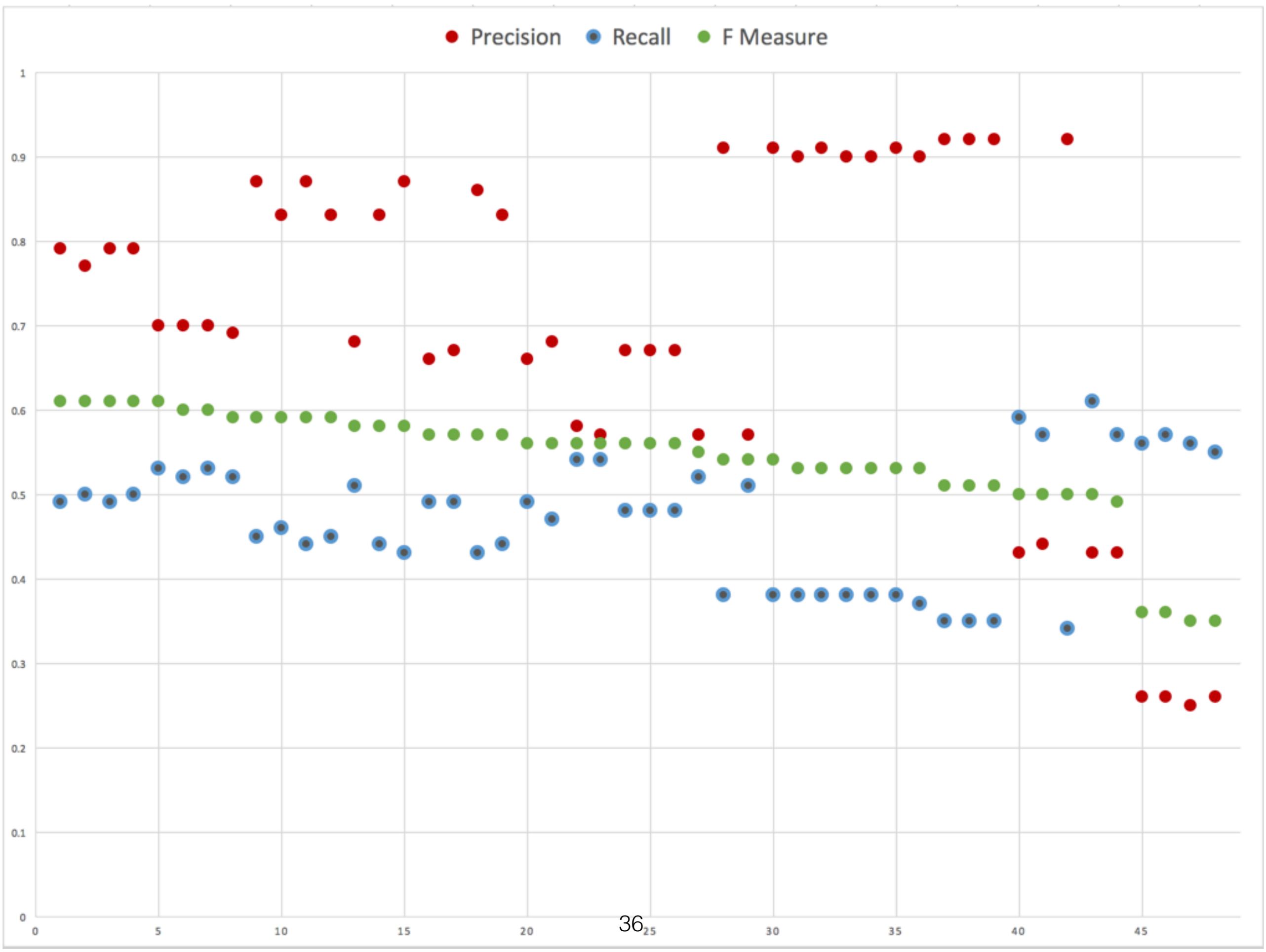


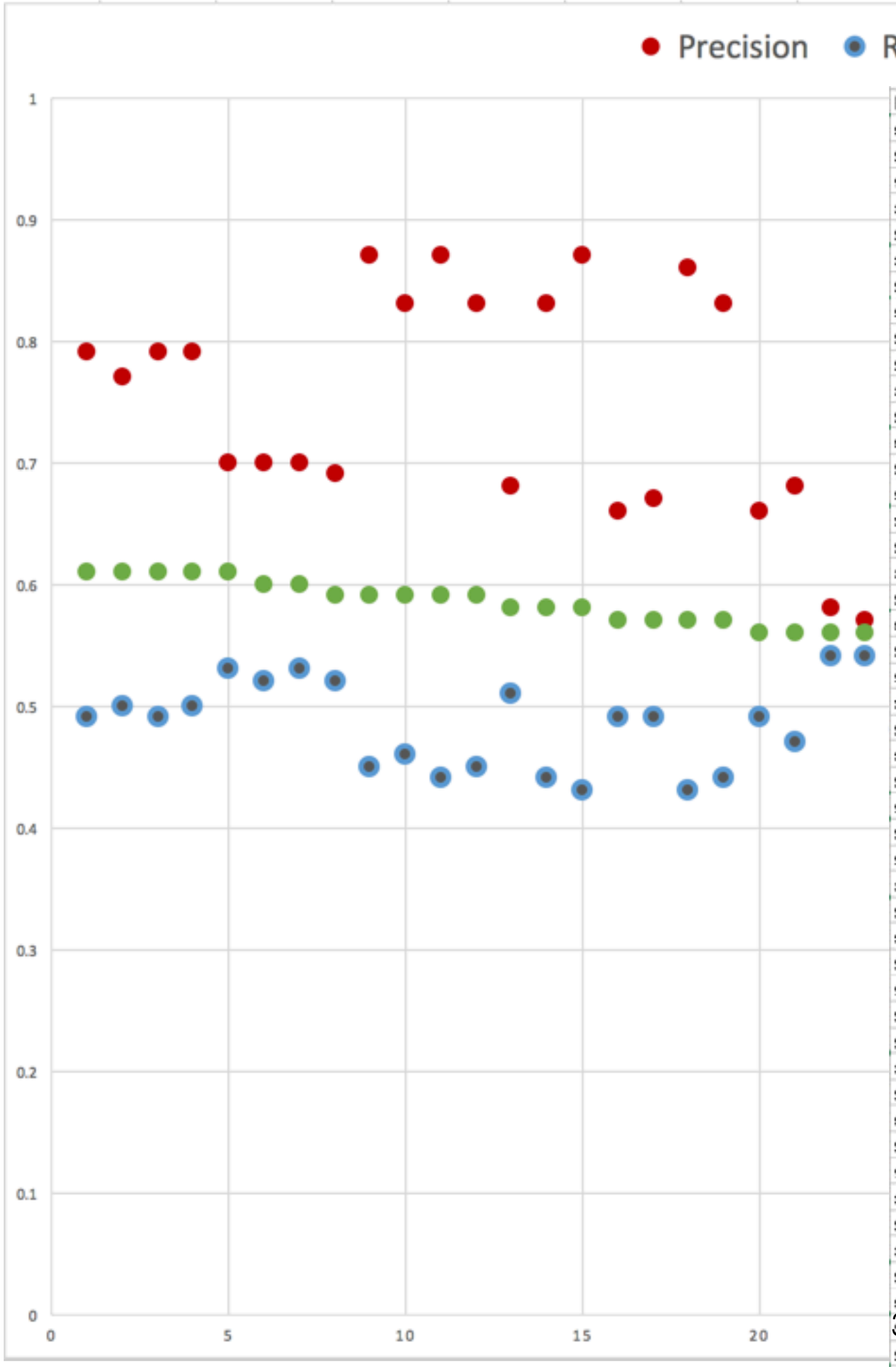
DataSet	Keys	Distance	Family Si	Merge Di	Precision	Recall	F Measur
skye	NamesOriginal	7	20	3	0.87	0.83	0.85
skye	NamesOriginal	16	10	3	0.89	0.8	0.84
skye	NamesOriginal	16	20	3	0.88	0.79	0.83
skye	NamesOriginal	3	10	3	0.89	0.77	0.83
skye	NamesOriginal	3	20	3	0.88	0.78	0.83
skye	NamesOriginal	7	10	3	0.89	0.78	0.83
skye	NamesOriginal	7	5	3	0.92	0.75	0.83
skye	NamesOriginal	9	10	3	0.89	0.77	0.83
skye	NamesOriginal	9	20	3	0.87	0.8	0.83
skye	NamesOriginal	9	5	3	0.92	0.75	0.83
skye	NamesOriginal	16	5	3	0.92	0.74	0.82
skye	NamesOriginal	3	5	3	0.92	0.74	0.82
skye	NamesOriginal	16	5	7	0.77	0.79	0.78
skye	NamesOriginal	7	5	7	0.79	0.77	0.78
skye	NamesOriginal	9	5	7	0.79	0.76	0.78
skye	NamesOriginal	3	5	7	0.79	0.76	0.77
skye	NamesOriginal	16	5	9	0.75	0.76	0.76
skye	NamesOriginal	9	5	9	0.75	0.76	0.76
skye	NamesOriginal	3	5	9	0.75	0.75	0.75
skye	NamesOriginal	7	5	9	0.75	0.75	0.75
skye	NamesOriginal	3	5	16	0.7	0.76	0.73
skye	NamesOriginal	7	5	16	0.69	0.77	0.73
skye	NamesOriginal	16	5	16	0.69	0.76	0.72
skye	NamesOriginal	9	5	16	0.7	0.75	0.72
skye	NamesOriginal	16	10	7	0.57	0.85	0.69
skye	NamesOriginal	3	10	7	0.58	0.85	0.69
skye	NamesOriginal	7	10	7	0.58	0.79	0.67
skye	NamesOriginal	9	10	7	0.57	0.79	0.67
skye	NamesOriginal	9	10	9	0.51	0.83	0.63
skye	NamesOriginal	16	10	9	0.51	0.8	0.62
skye	NamesOriginal	3	10	9	0.5	0.79	0.62
skye	NamesOriginal	7	10	9	0.51	0.77	0.61
skye	NamesOriginal	3	10	16	0.44	0.78	0.57
skye	NamesOriginal	16	10	16	0.44	0.78	0.56
skye	NamesOriginal	7	10	16	0.43	0.78	0.56
skye	NamesOriginal	9	10	16	0.45	0.76	0.56
skye	NamesOriginal	3	20	7	0.41	0.82	0.55
skye	NamesOriginal	16	20	7	0.41	0.79	0.54
skye	NamesOriginal	9	20	7	0.41	0.81	0.54
skye	NamesOriginal	7	20	7	0.39	0.82	0.53
skye	NamesOriginal	7	20	9	0.31	0.85	0.46
skye	NamesOriginal	9	20	9	0.32	0.85	0.46
skye	NamesOriginal	3	20	9	0.31	0.8	0.45
skye	NamesOriginal	16	20	9	0.31	0.79	0.44
skye	NamesOriginal	16	20	16	0.26	0.78	0.39
skye	NamesOriginal	7	20	16	0.26	0.78	0.39
skye	NamesOriginal	9	20	16	0.26	0.8	0.39
skye	NamesOriginal	3	20	16	0.25	0.77	0.38


Precision
Recall
F Measure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25
30
35
40
45

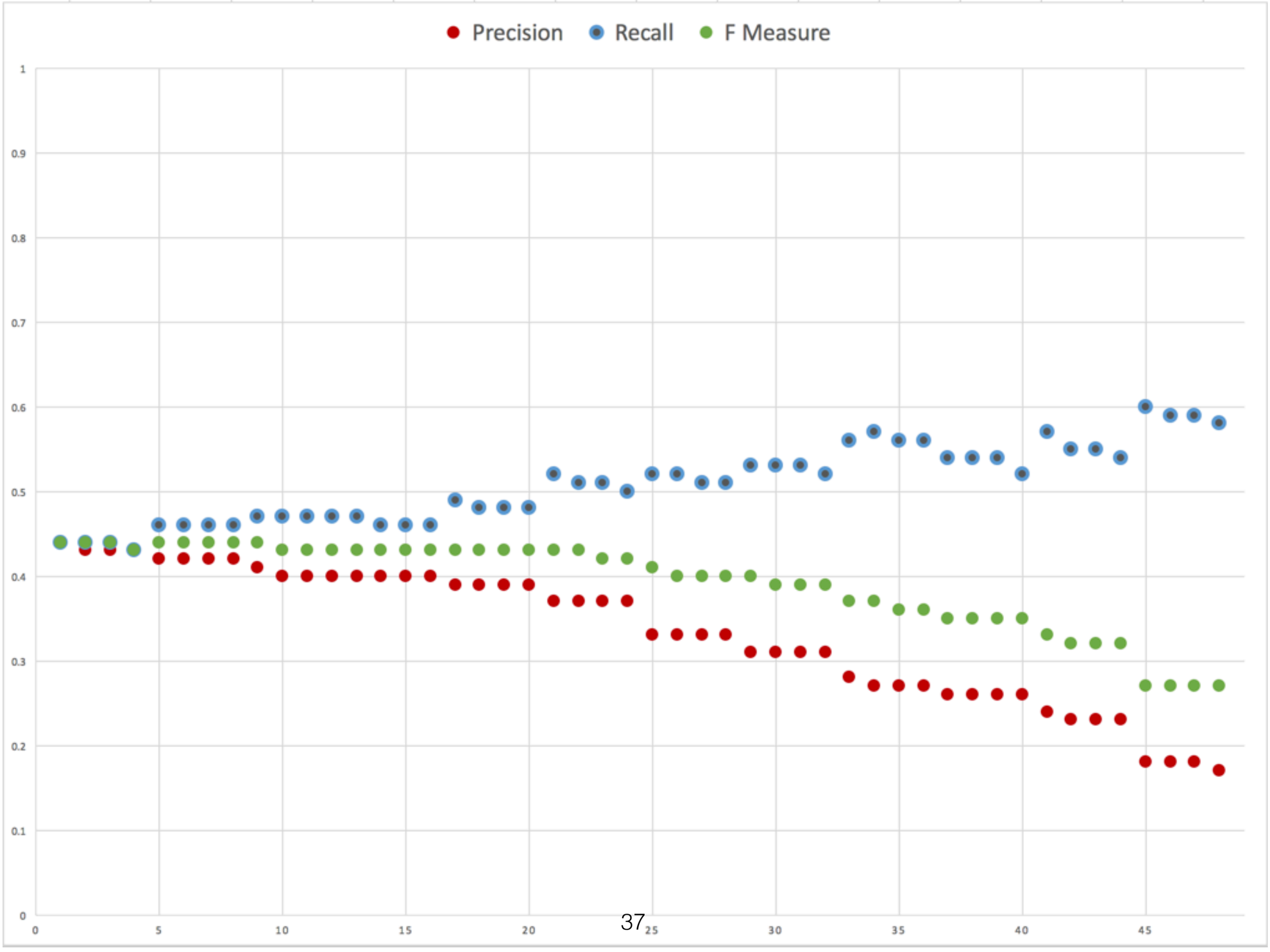


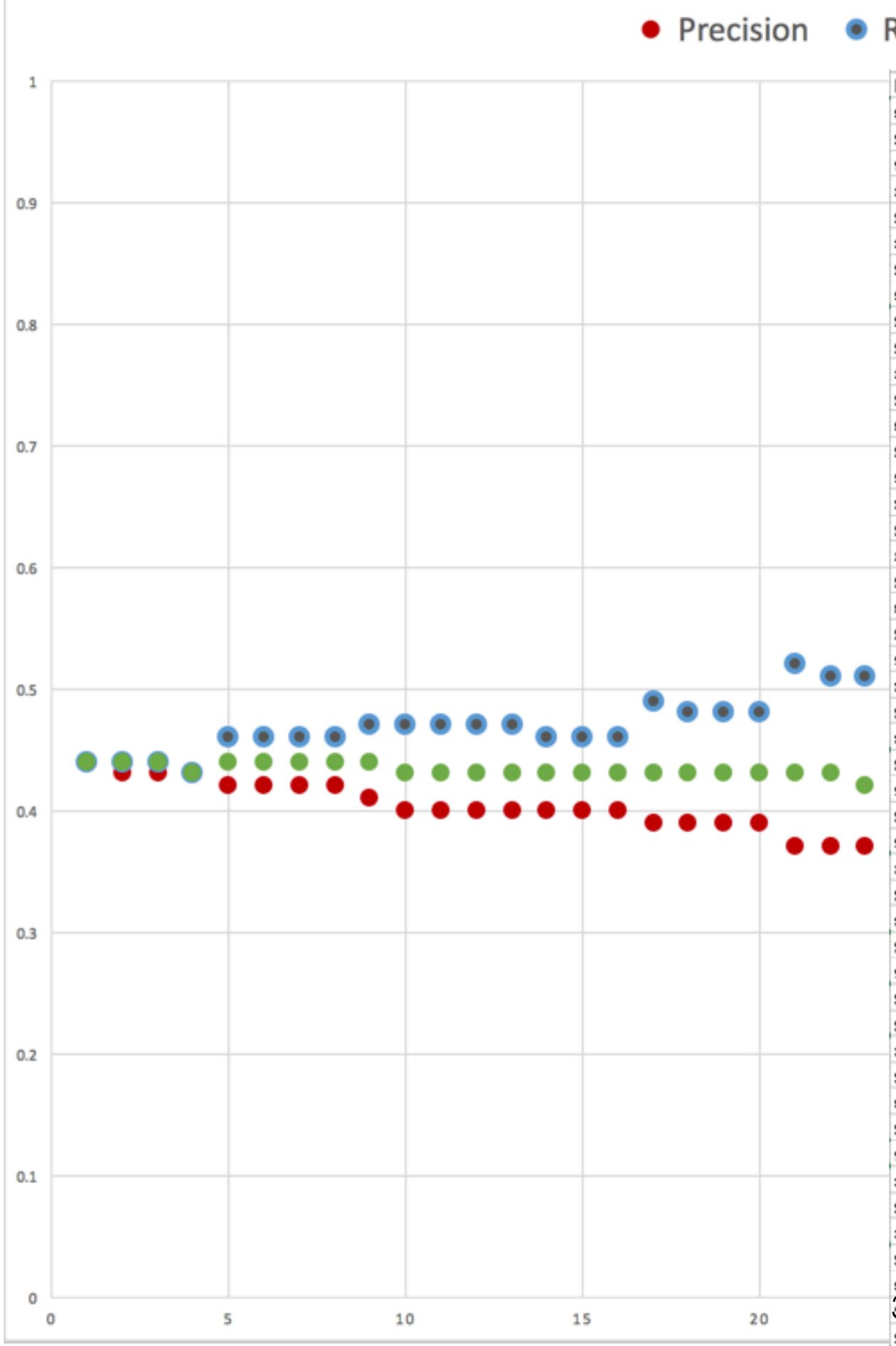
DataSet	Keys	Distance	Family Si	Merge Di	Precision	Recall	F Measur
skye	NamesClean	16	10	3	0.11	0.98	0.2
skye	NamesClean	16	20	3	0.11	0.98	0.2
skye	NamesClean	16	5	16	0.11	0.98	0.2
skye	NamesClean	16	5	3	0.11	0.98	0.2
skye	NamesClean	16	5	7	0.11	0.98	0.2
skye	NamesClean	16	5	9	0.11	0.98	0.2
skye	NamesClean	3	10	3	0.11	0.98	0.2
skye	NamesClean	3	20	3	0.11	0.98	0.2
skye	NamesClean	3	5	16	0.11	0.98	0.2
skye	NamesClean	3	5	3	0.11	0.98	0.2
skye	NamesClean	3	5	7	0.11	0.98	0.2
skye	NamesClean	3	5	9	0.11	0.98	0.2
skye	NamesClean	7	10	3	0.11	0.98	0.2
skye	NamesClean	7	20	3	0.11	0.98	0.2
skye	NamesClean	7	5	16	0.11	0.98	0.2
skye	NamesClean	7	5	3	0.11	0.98	0.2
skye	NamesClean	7	5	7	0.11	0.98	0.2
skye	NamesClean	7	5	9	0.11	0.98	0.2
skye	NamesClean	9	10	3	0.11	0.98	0.2
skye	NamesClean	9	20	3	0.11	0.98	0.2
skye	NamesClean	9	5	16	0.11	0.98	0.2
skye	NamesClean	9	5	3	0.11	0.98	0.2
skye	NamesClean	9	5	7	0.11	0.98	0.2
skye	NamesClean	9	5	9	0.11	0.98	0.2
skye	NamesClean	16	10	16	0.1	0.98	0.19
skye	NamesClean	16	10	7	0.11	0.98	0.19
skye	NamesClean	16	10	9	0.1	0.98	0.19
skye	NamesClean	3	10	16	0.1	0.98	0.19
skye	NamesClean	3	10	7	0.11	0.98	0.19
skye	NamesClean	3	10	9	0.1	0.98	0.19
skye	NamesClean	7	10	7	0.11	0.98	0.19
skye	NamesClean	7	10	9	0.1	0.98	0.19
skye	NamesClean	9	10	16	0.1	0.98	0.19
skye	NamesClean	9	10	7	0.11	0.98	0.19
skye	NamesClean	9	10	9	0.1	0.98	0.19
skye	NamesClean	16	20	7	0.1	0.98	0.18
skye	NamesClean	3	20	7	0.1	0.98	0.18
skye	NamesClean	7	10	16	0.1	0.98	0.18
skye	NamesClean	7	20	7	0.1	0.98	0.18
skye	NamesClean	9	20	7	0.1	0.98	0.18
skye	NamesClean	16	20	16	0.09	0.98	0.17
skye	NamesClean	16	20	9	0.09	0.98	0.17
skye	NamesClean	3	20	16	0.09	0.98	0.17
skye	NamesClean	3	20	9	0.09	0.98	0.17
skye	NamesClean	7	20	16	0.09	0.98	0.17
skye	NamesClean	7	20	9	0.09	0.98	0.17
skye	NamesClean	9	20	9	0.09	0.98	0.17
skye	NamesClean	9	20	16	0.09	0.98	0.16


Precision
Recall
F Measure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25
30
35
40
45



DataSet	Keys	Distance	Family Si	Merge Di	Precision	Recall	F Measur
skye	NamesMarriageOrig nal	16	10	7	0.79	0.49	0.61
skye	NamesMarriageOrig nal	3	10	7	0.77	0.5	0.61
skye	NamesMarriageOrig nal	7	10	7	0.79	0.49	0.61
skye	NamesMarriageOrig nal	9	10	7	0.79	0.5	0.61
skye	NamesMarriageOrig nal	9	10	9	0.7	0.53	0.61
skye	NamesMarriageOrig nal	3	10	9	0.7	0.52	0.6
skye	NamesMarriageOrig nal	7	10	9	0.7	0.53	0.6
skye	NamesMarriageOrig nal	16	10	9	0.69	0.52	0.59
skye	NamesMarriageOrig nal	16	5	7	0.87	0.45	0.59
skye	NamesMarriageOrig nal	3	5	9	0.83	0.46	0.59
skye	NamesMarriageOrig nal	7	5	7	0.87	0.44	0.59
skye	NamesMarriageOrig nal	7	5	9	0.83	0.45	0.59
skye	NamesMarriageOrig nal	16	20	7	0.68	0.51	0.58
skye	NamesMarriageOrig nal	16	5	9	0.83	0.44	0.58
skye	NamesMarriageOrig nal	3	5	7	0.87	0.43	0.58
skye	NamesMarriageOrig nal	3	5	16	0.66	0.49	0.57
skye	NamesMarriageOrig nal	7	20	7	0.67	0.49	0.57
skye	NamesMarriageOrig nal	9	5	7	0.86	0.43	0.57
skye	NamesMarriageOrig nal	9	5	9	0.83	0.44	0.57
skye	NamesMarriageOrig nal	16	5	16	0.66	0.49	0.56
skye	NamesMarriageOrig nal	3	20	7	0.68	0.47	0.56
skye	NamesMarriageOrig nal	3	20	9	0.58	0.54	0.56
skye	NamesMarriageOrig nal	7	20	9	0.57	0.54	0.56
skye	NamesMarriageOrig nal	7	5	16	0.67	0.48	0.56
skye	NamesMarriageOrig nal	9	20	7	0.67	0.48	0.56
skye	NamesMarriageOrig nal	9	5	16	0.67	0.48	0.56
skye	NamesMarriageOrig nal	9	20	9	0.57	0.52	0.55
skye	NamesMarriageOrig nal	16	10	3	0.91	0.38	0.54
skye	NamesMarriageOrig nal	16	20	9	0.57	0.51	0.54
skye	NamesMarriageOrig nal	7	10	3	0.91	0.38	0.54
skye	NamesMarriageOrig nal	16	20	3	0.9	0.38	0.53
skye	NamesMarriageOrig nal	3	10	3	0.91	0.38	0.53
skye	NamesMarriageOrig nal	3	20	3	0.9	0.38	0.53
skye	NamesMarriageOrig nal	7	20	3	0.9	0.38	0.53
skye	NamesMarriageOrig nal	9	10	3	0.91	0.38	0.53
skye	NamesMarriageOrig nal	9	20	3	0.9	0.37	0.53
skye	NamesMarriageOrig nal	16	5	3	0.92	0.35	0.51
skye	NamesMarriageOrig nal	7	5	3	0.92	0.35	0.51
skye	NamesMarriageOrig nal	9	5	3	0.92	0.35	0.51
skye	NamesMarriageOrig nal	16	10	16	0.43	0.59	0.5
skye	NamesMarriageOrig nal	3	10	16	0.44	0.57	0.5
skye	NamesMarriageOrig nal	3	5	3	0.92	0.34	0.5
skye	NamesMarriageOrig nal	9	10	16	0.43	0.61	0.5
skye	NamesMarriageOrig nal	7	10	16	0.43	0.57	0.49
skye	NamesMarriageOrig nal	3	20	16	0.26	0.56	0.36
skye	NamesMarriageOrig nal	7	20	16	0.26	0.57	0.36
skye	NamesMarriageOrig nal	16	20	16	0.25	0.56	0.35
skye	NamesMarriageOrig nal	9	20	16	0.26	0.55	0.35


Precision
Recall
F Measure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25
30
35
40
45



DataSet	Keys	Distance	Family Si	Merge Di	Precision	Recall	F Measur
skye	NamesMarriageClean	9	5	3	0.44	0.44	0.44
skye	NamesMarriageClean	16	5	3	0.43	0.44	0.44
skye	NamesMarriageClean	7	5	3	0.43	0.44	0.44
skye	NamesMarriageClean	3	5	3	0.43	0.43	0.43
skye	NamesMarriageClean	16	10	3	0.42	0.46	0.44
skye	NamesMarriageClean	3	10	3	0.42	0.46	0.44
skye	NamesMarriageClean	7	10	3	0.42	0.46	0.44
skye	NamesMarriageClean	9	10	3	0.42	0.46	0.44
skye	NamesMarriageClean	9	20	3	0.41	0.47	0.44
skye	NamesMarriageClean	16	5	7	0.4	0.47	0.43
skye	NamesMarriageClean	3	5	7	0.4	0.47	0.43
skye	NamesMarriageClean	7	5	7	0.4	0.47	0.43
skye	NamesMarriageClean	9	5	7	0.4	0.47	0.43
skye	NamesMarriageClean	16	20	3	0.4	0.46	0.43
skye	NamesMarriageClean	3	20	3	0.4	0.46	0.43
skye	NamesMarriageClean	7	20	3	0.4	0.46	0.43
skye	NamesMarriageClean	7	5	9	0.39	0.49	0.43
skye	NamesMarriageClean	16	5	9	0.39	0.48	0.43
skye	NamesMarriageClean	3	5	9	0.39	0.48	0.43
skye	NamesMarriageClean	9	5	9	0.39	0.48	0.43
skye	NamesMarriageClean	9	5	16	0.37	0.52	0.43
skye	NamesMarriageClean	16	5	16	0.37	0.51	0.43
skye	NamesMarriageClean	7	5	16	0.37	0.51	0.42
skye	NamesMarriageClean	3	5	16	0.37	0.5	0.42
skye	NamesMarriageClean	7	10	7	0.33	0.52	0.41
skye	NamesMarriageClean	3	10	7	0.33	0.52	0.4
skye	NamesMarriageClean	16	10	7	0.33	0.51	0.4
skye	NamesMarriageClean	9	10	7	0.33	0.51	0.4
skye	NamesMarriageClean	3	10	9	0.31	0.53	0.4
skye	NamesMarriageClean	16	10	9	0.31	0.53	0.39
skye	NamesMarriageClean	7	10	9	0.31	0.53	0.39
skye	NamesMarriageClean	9	10	9	0.31	0.52	0.39
skye	NamesMarriageClean	16	10	16	0.28	0.56	0.37
skye	NamesMarriageClean	9	10	16	0.27	0.57	0.37
skye	NamesMarriageClean	3	10	16	0.27	0.56	0.36
skye	NamesMarriageClean	7	10	16	0.27	0.56	0.36
skye	NamesMarriageClean	16	20	7	0.26	0.54	0.35
skye	NamesMarriageClean	3	20	7	0.26	0.54	0.35
skye	NamesMarriageClean	9	20	7	0.26	0.54	0.35
skye	NamesMarriageClean	7	20	7	0.26	0.52	0.35
skye	NamesMarriageClean	3	20	9	0.24	0.57	0.33
skye	NamesMarriageClean	16	20	9	0.23	0.55	0.32
skye	NamesMarriageClean	7	20	9	0.23	0.55	0.32
skye	NamesMarriageClean	9	20	9	0.23	0.54	0.32
skye	NamesMarriageClean	9	20	16	0.18	0.6	0.27
skye	NamesMarriageClean	16	20	16	0.18	0.59	0.27
skye	NamesMarriageClean	3	20	16	0.18	0.59	0.27
skye	NamesMarriageClean	7	20	16	0.17	0.58	0.27

Discussion

- We are exploring
  - CFU-based linkage
  - The M-Tree data structure
- Some preliminary results
- Soon to be extended to end-to-end linkage

Request for Comments

- This work is exploratory
- Different approaches: CFUs & M-Tree
- Alternatives to F-measure? AUC?
- What values should we expect for precision/recall?
- "Golden linked data" may not be golden
- Other string distance measures. Dates?
- Commonality of names (discriminatory power)
- Scalability